Citrate Synthase from the Hyperthermophilic Archaeon, Pyrococcus furiosus

Muir, Jacqueline M.

Award date:
1995

Awarding institution:
University of Bath

Citrate Synthase from the Hyperthermophilic Archaeon, *Pyrococcus furiosus*

submitted by Jacqueline M. Muir

for the degree of PhD

of the University of Bath

1995

Copyright

J. M. Muir

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To Duncan

ABSTRACT

A comparative approach has been used to investigate the molecular basis of protein thermostability using the central metabolic enzyme, citrate synthase, as a model protein. The ultimate goal of this research programme is a structural comparison of the citrate synthases from a mesophilic, moderately thermophilic and hyperthermophilic source. In pursuit of this goal, the work reported in this thesis describes the purification of citrate synthase from the hyperthermophilic Archaeon, *Pyrococcus furiosus*, and the cloning, sequencing and expression of the citrate synthase gene.

The gene was amplified in two parts from *P. furiosus* genomic DNA using the polymerase chain reaction (PCR). Amino acid sequence information from the purified enzyme and from a multiple sequence alignment of known citrate synthase sequences was used to design the oligonucleotide primers. The remainder of the gene was amplified from circularised genomic fragments by inverse PCR. The citrate synthase coding region was then expressed in *Escherichia coli* JM105 which allowed the production of adequate quantities of functional enzyme for thermostability and crystallographic analyses. The recombinant enzyme exhibited 50 % residual activity after an incubation of 20 min at 100°C. Preliminary crystallographic analyses demonstrated that the recombinant enzyme crystallised in a tetragonal space group. In addition, the primary sequence of the *P. furiosus* citrate synthase was homology-modelled to the crystal structure of the citrate synthase of the moderately thermophilic Archaeon, *Thermoplasma acidophilum*. Putative thermostabilising features were then predicted by assimilating the information provided by the homology model, a compositional analysis and a structurally-based sequence alignment of the *P. furiosus*, *T. acidophilum* and pig citrate synthase sequences. The primary sequence was also used, with other citrate synthase sequences, in a protein-derived phylogenetic analysis.

Acknowledgements

I would like to thank my supervisors Mike Danson, David Hough, and from Zeneca Bio Products, David Byrom, for their help and advice throughout this project. In addition, I would like to thank Richard Sharp and Neil Raven (Centre for Applied Microbiological Research, Porton Down) for providing *Pyrococcus furiosus* cell paste and DNA, Janice Young (Zeneca Pharmaceuticals, Macclesfield) for protein micro-sequencing and Rupert Russell and Garry Taylor for collaborations involving homology-modelling and crystallographic analyses. Thanks also to the members of the 'end' and 'middle' labs (past and present) for making the last three years enjoyable.

This work was funded by SERC (BBSRC) and Zeneca Bio Products, Billingham.

TABLE OF CONTENTS

Abbreviations

A_{280}	absorbance at 280 nm
bp	base pairs
BSA	bovine serum albumin
cDPG	cyclic 2, 3-diphosphoglycerate
CoA	coenzyme A
CS	citrate synthase
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
dGTP	deoxyguanosine 5'-triphosphate
DIG	digoxygenin
DIP	di-myo-inositol-1,1'-phosphate
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
dTTP	deoxythymidine 5'-triphosphate
EDTA	(disodium) ethylenediamine tetraacetate
EF	elongation factor
EPPS	N-[2-hydroxyethyl]piperazine-N'-[3-propane sulphonic acid]
$FAD(H_2)$	flavin adenine dinucleotide
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
IPTG	isopropyl-β-D-thiogalactoside
kb	kilobase pairs
K_d	distribution coefficient
kDa	kilodaltons
lacz	β-galactosidase gene
M_r	relative molecular mass

NAD(H)	nicotinic acid adenine dinucleotide
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenyl methyl sulphonyl fluoride
PVDF	polyvinylidene difluoride
rRNA	ribosomal ribonucleic acid
RNAse	ribonuclease A
S^0	elemental sulphur
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SSC buffer (1x)	0.15 M NaCl, 0.015 M trisodium citrate
TAE	40 mM Tris-acetate, 1mM EDTA
TBE	90 mM Tris-borate, 2 mM EDTA
TCA	trichloroacetic acid
TE buffer (1x)	10 mM Tris-HCl pH 8.0, 1 mM EDTA
TEMED	N,N,N',N'-tetramethylethylene diamine
T_m	Melting temperature
Tris	tris-(hydroxymethyl)-methylamine
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER 1

INTRODUCTION

This thesis concerns the cloning, sequencing and expression of the gene encoding citrate synthase from the hyperthermophilic Archaeon, *Pyrococcus furiosus*. Hence, this introduction will concentrate on the structure and mechanism of citrate synthase, the classification of the Archaea and the characteristics of *P. furiosus*. In addition, a review of recent studies of protein stability will be included.

1.1 CITRATE SYNTHASE

1.1.1 Citrate synthase as a model for investigating protein thermostability

Citrate synthase catalyses the condensation of acetyl-Coenzyme A (acetyl-CoA) and oxaloacetate to form citrate and free CoA (Figure 1.1). By this, it effects the entry of carbon into the citric acid cycle (Figure 1.2) and is regulated according to the energetic or biosynthetic requirements of the cell. The enzyme of eukaryal, archaeal and Gram-positive bacterial sources is a homodimer that is isosterically regulated by ATP, whereas the enzyme from most Gram-negative sources is a homohexamer, allosterically regulated by NADH [Danson, 1988]. The known exceptions are the dimeric, ATP-regulated citrate synthases of *Coxiella burnetii* and *Rickettsia prowazekii*; this may be a function of their intracellular parasitic lifestyle. The citrate synthase of Gram-negative organisms with a facultatively anaerobic metabolism is also regulated by α-ketoglutarate, reflecting the biosynthetic role of the citric acid cycle in these organisms.

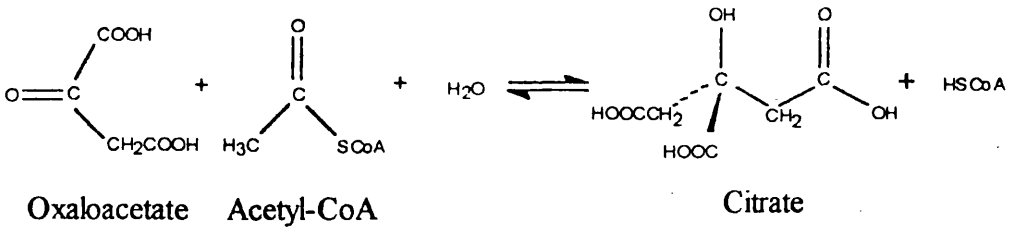


Figure 1.1 The reaction catalysed by citrate synthase

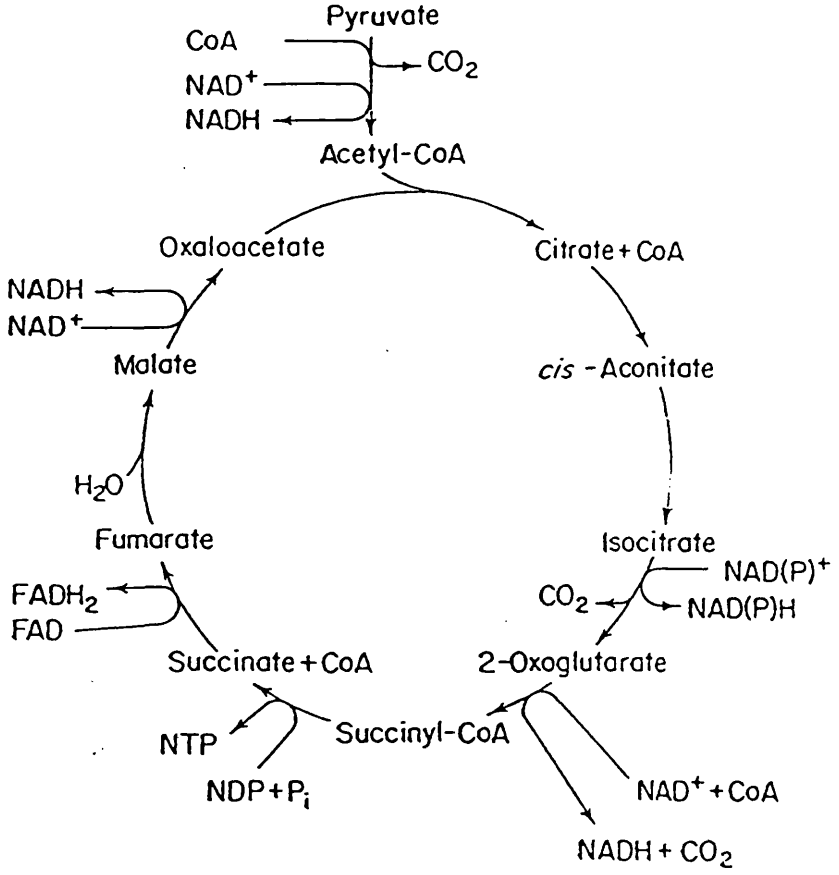


Figure 1.2 The citric acid cycle

The primary sequence of eighteen citrate synthases has been elucidated (Table 1.1). Seven of these are from eukaryal sources, ten from bacterial sources (six Gram-positive and four Gram-negative organisms) and one from an archaeal species. Of these, both *Saccharomyces cerevisiae* and *Bacillus subtilis* possess two citrate synthases [Rosekrantz *et al.*, 1986; Jin and Sonenshein, 1994]. The former has a mitochondrial and a glyoxysomal form of the enzyme whereas the Gram-positive Bacterium has a predominant form, CitZ, and a second enzyme, CitA, the gene for which can be disrupted without any obvious abnormal effect on cell metabolism. The existence of two citrate synthases has also been demonstrated in two Gram-negative Bacteria, *Pseudomonas aeruginosa* [Anderson *et al.*, 1993] and *Escherichia coli* [Patton *et al.*, 1993]; both species possess a second, dimeric citrate synthase (see also Section 3.4).

Source of citrate synthase sequence	Reference
Pig	Evans *et al.*, 1988
Chicken	Karpusas *et al.*, 1990
Caenorhabditis (*C.*) *elegans*	Swiss-Prot Data Bank - P34575
Yeast (mitochondrial form)	Rosenkrantz *et al.*, 1986
Yeast (glyoxysomal form)	Rosenkrantz *et al.*, 1986
Neurospora (*N.*) *crassa*	EMBL Data Bank - P34085
Tetrahymena (*Te.*) *thermophila*	Numata *et al.*, 1991
Acinetobacter (*A.*) *anitratum*	Donald *et al.*, 1987
Pseudomonas (*Ps.*) *aeruginosa*	Donald *et al.*, 1989
Escherichia (*E.*) *coli*	Ner *et al.*, 1983
Acetobacter (*Ac.*) *acetii*	Fukaya *et al.*, 1990
Coxiella (*Co.*) *burnetii*	Heinzen *et al.*, 1991
Rickettsia (*R.*) *prowazekii*	Wood *et al.*, 1987
Bacillus (*B.*) *subtilis* (CitZ)	Jin and Sonenshein, 1994
Bacillus (*B.*) *subtilis* (CitA)	Jin and Sonenshein, 1994
Bacillus (*B.*) *coagulans* (strain C4)	Schendel *et al.*, 1992
Mycobacterium (*M.*) *smegmatis*	David *et al.*, 1991
Thermoplasma (*T.*) *acidophilum*	Sutherland *et al.*, 1990

Table 1.1 Sources of known citrate synthase sequences

(The citrate synthase sequences of *C. elegans* and *N. crassa* are referenced only by their Data Bank Accession numbers.)

Citrate synthase is an ideal model protein for an investigation into the molecular basis of protein thermostability. It is present in virtually all organisms and has maintained the same function throughout evolution. There is a substantial data base of sequence information for this enzyme and, more importantly, the crystal structures are known for the enzyme from pig heart, chicken heart and the thermophilic Archaeon, *Thermoplasma acidophilum*, as described below. In addition, the activity of the enzyme is measured by a simple spectrophotometric assay, facilitating enzyme purification and thermostability analyses.

1.1.2 The structure of the citrate synthase

The elucidation of the structure of the pig heart citrate synthase was achieved in 1982 by Remington and co-workers [Remington *et al.*, 1982]. The pig heart enzyme was crystallised, and the structure solved in two forms, designated 'open' and 'closed' (Figure 1.3). The secondary structure of the enzyme is predominantly α-helical, each subunit consisting of a large (15 α-helices) and a small (5 α-helices) domain. The overall difference between the two forms can be approximated by a 19° rotation of the small domain relative to the large, although subtle rearrangement of internal side chains also occurs. The resulting exclusion of substrates and active site residues from the bulk solvent is essential for the proposed catalytic mechanism, detailed in Section 1.1.3.

The 20 α-helices of each subunit of the pig heart citrate synthase are designated A-T; the helices of the second subunit are designated A'-T'. As shown in Figure 1.4, a schematic view of the citrate synthase monomer, there is a small region of β-sheet and long loop regions between the α-helices. The dimer is formed by a two-fold rotational symmetry operation such that eight α-helices are located at the subunit interface in four antiparallel pairs (Helices FF', GG', LL' and MM'). These are orientated perpendicularly to the axis of rotation and comprise the bulk of the subunit interface. Each citrate synthase dimer has two active sites, located in a cleft between the large and

the small domains of each subunit. Residues from both subunits contribute to each active site although the sites are independent and no co-operativity of binding or catalysis has been observed. The two terminal α-helices, A and T, that appear to be distant from the main bulk of the subunit in Figure 1.4, actually lie in grooves upon the surface of the second subunit in the dimer. Incidentally, α-helix A and the long loop region connecting this helix to α-helix B, constitute the amino-terminal extension that is absent in citrate synthases from archaeal and Gram-positive sources studied to date (Section 3.4). Many of the α-helices of the pig heart enzyme appear kinked or curved over a large angle; this has been attributed either to the presence of a Pro residue at an internal position in the α-helix, or by the inclusion of water molecules in the α-helix hydrogen bonding pattern.

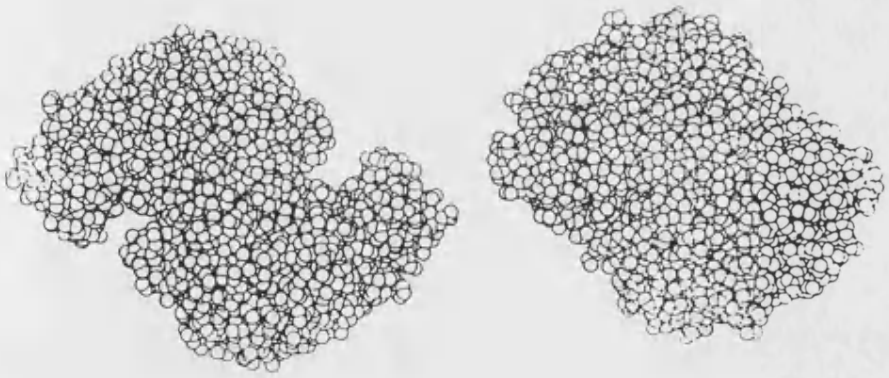


Figure 1.3 Space-filling drawing of the open and closed forms of the pig heart citrate synthase.

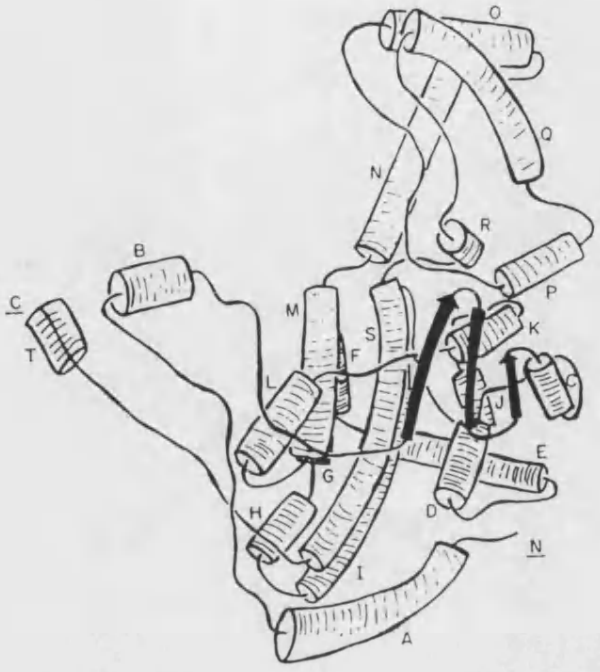


Figure 1.4 Schematic view of the monomer of the pig heart citrate synthase looking down the two-fold axis of symmetry. Helices N, O, P, Q and R constitute the small domain whereas Helices A-M, S and T constitute the large domain. N and C refer to the amino- and carboxyl-termini. (Reproduced from Remington, 1982)

The elucidation of the crystal structure of the enzyme from pig or chicken heart complexed with substrates, substrate analogues, products or putative transition state analogues has been essential in elucidating the substrate binding residues [Karpusas *et al.*, 1990, 1991; Remington, 1992]. In conjunction with mutagenesis experiments involving putative active site residues [Alter *et al.*, 1990; Zhi *et al.*, 1991], a catalytic mechanism has been suggested for the pig heart enzyme [Remington, 1992].

1.1.3 The catalytic mechanism of citrate synthase

Oxaloacetate is ligated in the active site, located between the large and small domains, by three Arg residues and three His residues such that the molecule is completely surrounded by polar or obligatory-charged groups. The C3 carbon of oxaloacetate is surrounded by positive, electron-withdrawing groups which bestow a partial positive charge on this carbon and render it susceptible to nucleophilic attack. The orientation of oxaloacetate within the active site ensures the stereospecificity of the reaction by exposing only one side of the substrate molecule to acetyl-CoA. Acetyl-CoA has been shown to bind to the enzyme through a narrow cleft such that only the acetyl group and the terminal sulphur atom can access oxaloacetate.

The reaction catalysed by citrate synthase is a stereospecific condensation of oxaloacetate and acetyl-CoA, precipitated by the nucleophilic attack on the C3 carbon of oxaloacetate by the methyl carbon of acetyl-CoA. The reaction can be subdivided into three steps: enolisation of acetyl-CoA, condensation of acetyl-CoA and oxaloacetate and thioester hydrolysis to release products. The enolisation of acetyl-CoA is achieved by concerted acid-base catalysis whereby His274 (of the pig enzyme) protonates the carbonyl oxygen of acetyl-CoA while, simultaneously, Asp375 extracts a proton from the adjacent methyl moiety (Figure 1.5a). Thus, the methyl moiety is prepared to attack, nucleophilically, the C3 carbon of oxaloacetate. A second acid-base catalysis has been proposed to effect such an attack. His320 (of the pig enzyme) is thought to protonate the carbonyl oxygen of oxaloacetate while His274 is reprotonated in preparation for the next enolisation (Figure 1.5a). The mechanism of product release by thioester hydrolysis is not fully understood, but experiments involving the creation of active site mutants of pig citrate synthase have implied that Asp375 and His274 are essential [Alter *et al.*, 1990]. Two schemes have been suggested. Asp375 may protonate the sulphur of acetyl-CoA which is then susceptible to hydrolysis by an activated water (as depicted in Figure 1.5b), or Asp375 and the reaction intermediate may form an anhydride which is then hydrolysed by an activated water.

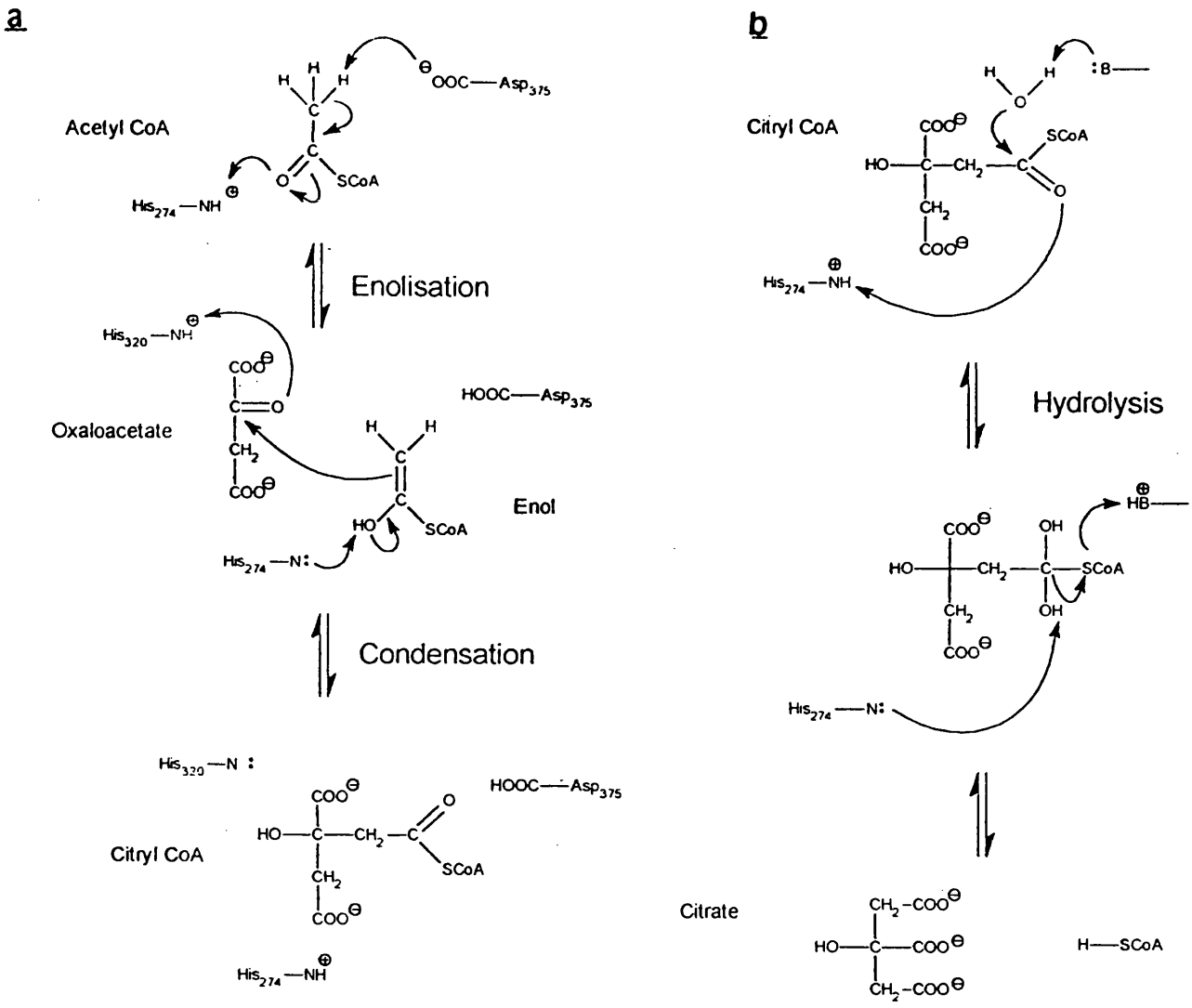


Figure 1.5 The catalytic mechanism of citrate synthase

(a) The enolisation of acetyl-CoA and the condensation of acetyl-CoA and oxaloacetate

(b) A suggested mechanism for thioester hydrolysis - protonation of the acetyl-CoA sulphur leads to hydrolysis via an activated water molecule.

1.1.4 The crystal structure of the *Thermoplasma acidophilum* citrate synthase

Thermoplasma acidophilum is a moderately thermophilic Archaeon that has an optimal growth rate at 55°C. The crystal structure of the enzyme from pig heart was used as a search model for the elucidation of the *T. acidophilum* crystal structure by the method of molecular replacement [Russell *et al.*, 1994]. This illustrates the structural conservation of this enzyme between the eukaryal and the archaeal organisms, even though the sequence identity between the two enzymes was just 20 %. Such structural conservation allows comparison of the enzyme from the moderately thermophilic source with that of the mesophilic source and has highlighted a number of putative, thermostabilising features.

The *T. acidophilum* enzyme crystallised in the 'open' form of the enzyme and thus, allowed direct comparison with the equivalent structure of the pig enzyme [Russell *et al.* 1994]. The archaeal enzyme comprised just 16 α-helices; the helices equivalent to A and H of the pig heart enzyme (Figure 1.4) are absent. Due to diffuse electron density at the amino- and carboxyl-termini of the subunit polypeptide, the positions of α-helices equivalent to B and T of the pig heart enzyme could not be defined.

A number of differences were noted between the two structures, most strikingly, the absence of both the amino-terminal α-helix and the internal α-helix, H, and the presence of shorter loops between α-helices in the thermostable enzyme. Thus, the *T. acidophilum* enzyme is smaller and more compact. This has also been demonstrated by a difference in the orientation of the small and the large domains of each subunit which may be considered as a 'closing' of the two domains in the thermostable enzyme. This phenomenon has been observed in two other comparative systems, as detailed in Section 1.4.3. As in the pig enzyme, the *T. acidophilum* structure has four pairs of antiparallel α-helices at the subunit interface; one of these α-helices is particularly Ala-rich which may be enhancing both the α-helix stability (see

Section 1.4.2) and the hydrophobic contacts between the two subunits. The contribution of internal hydrophobic packing to the stability of the folded conformation is discussed in Section 1.4.2 [Dill, 1990; Jaenicke, 1991]. In agreement with these ideas, the moderately thermostable citrate synthase possessed fewer internal cavities than the pig heart enzyme, and those remaining were smaller in size. An increase in the number of interactions between aromatic side chains was also noted in the small domain of each subunit.

Thus, the overall impression of the moderately-thermostable citrate synthase is of a compact, tightly folded enzyme with regions of increased hydrophobic packing and amino acid exchanges to residues of greater α-helical propensity. Studies are underway to test these hypotheses by using site-directed mutagenesis to exchange amino acids in the pig heart enzyme (M. M^{c}Cormack, University of Bath). To complement and extend this work, the aim of the work reported in this thesis is to investigate the enzyme from a hyperthermophilic Archaeon, *Pyrococcus furiosus*, with a view to a structural comparison of the enzyme from a mesophilic, moderately thermophilic and hyperthermophilic source. However, prior to detailing the aims of this individual project it is necessary to introduce the organism and its phylogenetic origin and to discuss in detail the progress made in other systems towards assessing the stability of the folded conformation.

1.2 DETERMINATION OF THE ARCHAEA AS A MONOPHYLETIC GROUP

1.2.1 Methods of phylogenetic inference

The information contained within amino acid or nucleic acid sequences has effectively increased the resolution at which the evolutionary relatedness of organisms can be determined. In other words, the classification of organisms, especially microbes, no longer relies on macroscopic, phenotypic characteristics to imply an evolutionary relationship. Low resolution classification has been revolutionised by a number of statistical methods that attempt to evaluate the evolutionary distance between species, and thus provide a natural scheme of systematics. It must be noted, that each method is inferring the evolutionary history of a species by assuming a model of the evolutionary process. Thus, the information must be interpreted as the maximum possibility that, using a particular model of evolution, a particular relationship occurs.

The central model used in all phylogenetic methods is one of evolution by random mutation. In many cases, a molecular clock is assumed in the analysis thereby implying that the rate of change is constant for a particular protein, gene or part of the genome. This can be explained by the conservation of biological function, encouraging the accumulation of neutral or selective mutations. One of the difficulties in assuming a particular model of the evolutionary process is assessing the weighting required for neutral, selective or deleterious mutations and how to evaluate the reversion of a character to one that is present in a distantly related species. However, the following methods of analysis have been developed to compute the evolutionary history of an organism from molecular sequence data [reviewed by Felsenstein, 1988].

Parsimony methods compare a set of phylogenetically informative sequences and propose a phylogenetic tree based on the minimum number of mutations required to relate each sequence to the others. Distance matrix methods determine the distance between pairs of phylogenetically informative sequences and construct the most probable tree such that the branch distance between two species represents the evolutionary distance between them. Likelihood methods use a statistical approach such that there is a probability that the applied data produce a certain tree given a model of the evolutionary process.

1.2.2 Consistency and resampling

The correctness of a particular tree can only be ascertained by applying certain criteria, in particular consistency. If a phylogenetic method is consistent then the addition of more data brings the tree topology closer to the true phylogenetic relationship. Felsenstein [1993] urges a statistical approach to interpreting phylogenies such that a computated, phylogenetic tree is considered a snapshot of the optimum of a distribution of branch lengths and node positions.

One procedure that is followed in many statistical analyses is a resampling procedure whereby the possibility of a small region of sequence greatly influencing the outcome is eliminated. In phylogenetic analyses, the bootstrap and jacknife resampling methods are most commonly used. The resampled data are then used to produce a tree topology which can be compared to the original and to the others produced by the resampling method.

1.2.3 Choice of molecular sequence

By performing a phylogenetic analysis, one is automatically inferring the evolutionary history of an organism by one gene or protein [Olsen and Woese, 1993]. Therefore, it is necessary that the gene or protein in question is of essential function and thereby fundamental to all extant organisms. The molecule must also be sufficiently informative to trace the molecular history of the species; enough sequence is required to recognise relatedness, yet useful information for phylogeny requires variable sites. For this reason, ribosomal RNA (rRNA) has emerged as the favoured molecular chronometer.

Protein sequence comparisons have an intrinsic advantage over nucleotide sequences in that the likelihood of random mutations leading to a non-related similarity in sequence is remote. However, phylogenies derived from protein sequences must account for conservative amino acid substitutions that occur frequently and have a minimal effect on the function or efficiency of the protein. In addition, it has been suggested [Olsen and Woese, 1993] that the protein composition may reflect the underlying nucleotide composition of the organism or the stabilisation of the protein to the environment of the organism. Therefore, a phylogenetic analysis of the available citrate synthase sequences must be considered with awareness of such limitations

1.2.4 Global classification using rRNA as a chronometer

In addition to the long-awaited, natural classification of micro-organisms, the use of molecular sequences as evolutionary markers has challenged current views as to the nature of global classification. Using 16S rRNA sequences, Woese and Fox [1977] disputed the phenotypically-derived, prokaryote-eukaryote divide. Using the parsimony method of phylogenetic inference, these researchers provided evidence that supported the long-standing hypotheses of many microbiologists such as Stanier and van Niel

[1962], that microbes are as diverse from each other as they are from eukaryotes as a whole. Further studies by Woese and co-workers [Woese *et al.*, 1990; Wheelis *et al.*, 1992] led to the definition of three primary domains of life: the Eukarya, the Bacteria and the Archaea (Figure 1.6).

Like any new evolutionary concept, this was not received without objection from supporters of both the prokaryotic-eukaryotic and the five kingdom classification schemes. One objector to the Woese classification system, James Lake [1988], used similar data but a different approach to produce a tree that did not demonstrate the Archaea to be a monophyletic unit. He argued that it was necessary to account for rate variances in adjacent branches of a proposed tree. Although some controversy continues, evidence continues to accumulate in support of the tree defined by Woese and co-workers [Woese *et al.*, 1990].

1.2.5 Rooting the tree

Aside from the controversy over the topology of the global phylogenetic tree, Iwabe and co-workers [Iwabe *et al.*, 1989] pursued ways to define the position of the root. This would predict the relatedness of extant organisms to the postulated universal ancestor or 'progenote'. When relating all organisms, no outgroup species is available. Therefore, the position of the root was defined using two pairs of genes that were known to have duplicated and diverged before the division into the primary domains of life. The duplicated genes selected by Iwabe and co-workers [1989] were those encoding the α and β subunits of F_1 ATPases and the elongation factors EF-1α and EF-2. By these two sets of duplicates alone, the tree was rooted (Figure 1.6) such that the Eukarya and Archaea were sister groups with the Bacteria more distantly related.

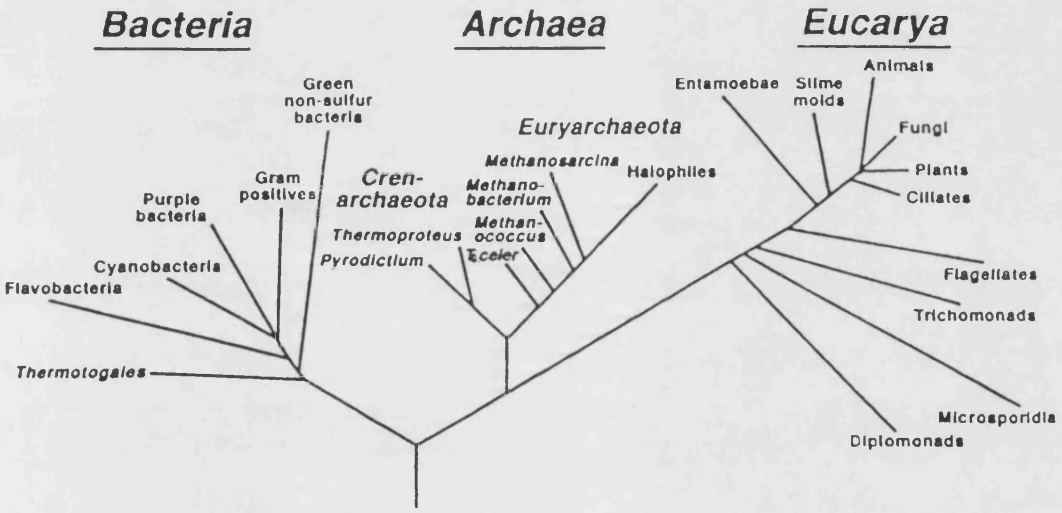


Figure 1.6 The universal phylogenetic tree, as determined by phylogenetic analysis of 16S rRNA sequences (reproduced from Olsen and Woese, [1993]). The tree was rooted as described by Iwabe *et al.*, [1989].

1.2.6 The archaeal tree topology - recent developments

The archaeal tree was shown by Woese [Woese *et al.*, 1990] to exhibit a bifurcation giving rise to two kingdoms, the Crenarchaeota and the Euryarchaeota. The Crenarchaeota is considered to be a monophenetic kingdom as it consists of just thermophilic organisms. The Euryarchaeota encompasses a more diverse range of phenotypes in that it comprises halophiles, methanogens and thermophiles. It must be noted, that the euryarchaeal hyperthermophiles are positioned nearest to the bifurcation point. A study by Kjems and co-workers [1992] using 23S rRNA supported the 16S rRNA analysis of Woese [Woese *et al.*, 1990] and, with particular relevance to this study, demonstrated the relatedness of *P. furiosus* to *Thermococcus* (*Tc.*) *celer* and the position of these species near to the root of the global, phylogenetic tree (Figure 1.6). The definition of *P. furiosus* as a Euryarchaeote is important as it is in the same kingdom as the moderate thermophile, *T. acidophilum*. Thus, comparison of sequence and structure between the citrate synthases of the moderate and extreme thermophile will be minimally biased by phylogenetically-derived differences.

The development of molecular techniques, in particular the polymerase chain reaction (PCR), has allowed the determination of the sequences of rRNA from organisms that have not yet been cultured [Barns *et al.*, 1994] This has demonstrated the diversity of archaeal species and has indicated that the divide between the Euryarchaeota and the Crenarchaeota may be less defined than originally thought. A recent study [Delong *et al.*, 1994], using similar techniques, has demonstrated a new archaeal habitat, that of the Antarctic ocean at -1.5°C.

1.2.7 Phylogeny and this project

The evolutionary relatedness of organisms will always be in question as even the most probable relationships must be declared as a function of the statistical methods and the underlying assumptions applied. Providing the phylogenetic analyses are of maximal consistency, analysis of different proteins, genes or regions of the genome should converge to a highly probable, evolutionary hypothesis. The description of *P. furiosus* as a hyperthermophilic Archaeon immediately indicates acceptance of the Archaea as a monophyletic unit. However, phylogenetic analysis of the citrate synthase sequence of this organism will contribute to a protein-derived phylogeny which may complement or dispute the universal tree topology. Therefore, while accepting the phylogenetic tree of Woese, it is now essential to describe the biochemical characteristics of the hyperthermophilic Archaeon, *Pyrococcus furiosus*.

1.3 *PYROCOCCUS FURIOSUS*

P. furiosus is a marine hyperthermophilic Archaeon, isolated in 1986 by Fiala and Stetter from geothermally-heated, marine sediments on Vulcano Island, Italy [Fiala and Stetter, 1986]. It is strictly anaerobic and has an optimal growth rate at 100°C. The cells are spherical, 0.8-2.5 μm in diameter; the base content of the genome is 38 % G+C. Only four other genera exhibit an optimal growth rate at, or above, 100°C; these are *Pyrobaculum* [Huber *et al.*, 1987], *Pyrodictium* [Stetter *et al.*, 1983], *Hyperthermus* [Zillig *et al.*, 1990] and *Methanopyrus* [Huber *et al.*, 1989]. Of the four non-methanogenic genera, only two species, *Pyrococcus furiosus* and *Pyrococcus abyssi* [Erauso *et al.*, 1993] can be cultured in the presence or absence of elemental sulphur (S^0). Members of the other genera obtain energy by reducing S^0 with either hydrogen (*Pyrodictium sp.*) or organic substrates (*Pyrobaculum sp.*, *Hyperthermus sp.*, *Pyrococcus woesei*).

The anaerobic metabolism of *P. furiosus* has become subject to debate. Adams and co-workers isolated an aldehyde oxidoreductase and its associated redox co-factor, ferredoxin from *P. furiosus* [Mukund and Adams, 1991]. Ferredoxin was also shown to donate electrons to a hydrogenase, resulting in the evolution of molecular hydrogen [Bryant and Adams, 1989]. In culture, this is inhibitory to the organism in the absence of S^0, inferring that the reduction of S^0 to H_2S is a mechanism by which the organism detoxifies its environment. If hydrogen is removed by, for example, flushing nitrogen through the culture vessel, *P. furiosus* can be cultured in the absence of S^0. A novel metabolic pathway, 'pyroglycolysis', that was proposed to be operating in this organism [Mukund and Adams, 1991], was outlined by Schäfer and Schönheit [1991, 1992]. The pathway, based on the non-phosphorylated Entner-Doudroff pathway found in some aerobic Archaea [Danson, 1988, 1993 and references therein], described the degradation of maltose via glucose and pyruvate to acetyl-CoA without the involvement of thermolabile co-factors such as NAD(P)H. The measurement of activity of some of the enzymes of the Emden-Meyerhoff pathway was attributed to the pathway operating in reverse, as a gluconeogenic pathway [Schäfer and Schönheit, 1993].

Recently, the 'missing' activities of enzymes of the glycolytic pathway were demonstrated by Kengen and co-workers [Kengen *et al.*, 1994]. Interestingly, the two sugar kinases, hexokinase and phosphofructokinase, were shown to be ADP-linked rather than ATP-linked which may be another example of co-factor adaptation in *P. furiosus*. As the activity of all enzymes of the Emden-Meyerhoff pathway had been measured, it was suggested that this was the major pathway of glucose oxidation in this organism.

During these studies of *P. furiosus* metabolism, a number of proteins have been purified including aldehyde-ferredoxin oxidoreductase [Mukund and Adams, 1991], pyruvate-ferredoxin oxidoreductase [Blamey and Adams, 1993] and hydrogenase [Bryant and Adams, 1989]. Other enzymes isolated from this organism include a protease [Blumentals *et al.*, 1990; Eggen *et al.*, 1990], amylase [Koch *et al.*, 1990], α-glucosidase [Constantino *et al.*, 1990], rubredoxin [Blake *et al.*, 1991], glutamate dehydrogenase [Consalvi *et al.*, 1991, Robb *et al.*, 1992] and DNA polymerase [Lundberg *et al.*, 1991]. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) has been isolated from the related species, *P. woesei* [Zwickl *et al.*, 1990]. Of these, the extracellular enzymes have been shown to be the most thermostable; for example, amylase exhibited a half-life of 2 h at 120°C [Koch *et al.*, 1990]. Intracellular enzymes such as pyruvate oxidoreductase and GAPDH have shorter half-lives at lower temperatures (0.3 h at 90°C and 0.7 h at 100°C, respectively). This has prompted the proposal by Hensel and co-workers that the intracellular environment of the organism provides extrinsic stabilising factors which may help the organism to survive over a wider range of growth temperatures [Hensel and König, 1988; Scholz *et al.*, 1992]. This is discussed further in Section 6.4.3.

The fermentative nature of this organism infers that the citric acid cycle is used solely for biosynthetic purposes; this may be reflected in the intracellular levels of citric acid cycle enzymes in this organism. However, as this organism is not dependent on S^0 for growth and the subsequent generation of fermenter-deteriorating H_2S is avoided, *P. furiosus* is becoming the representative, archaeal hyperthermophile. Therefore, as the organism was readily available, it was selected as the hyperthermophilic source for citrate synthase.

1.4 STRUCTURAL BASIS OF PROTEIN THERMOSTABILITY

1.4.1 Stability of the folded conformation

The native conformation of a globular protein is maintained by an internal architecture of non-covalent interactions, namely electrostatic bonds, hydrogen bonds, van der Waals interactions and hydrophobic interactions [reviewed by Dill, 1990; Jaenicke, 1991]. If these are disrupted by the application of heat or a chaotrophic agent such as urea, a more diffuse, unfolded conformation is favoured and the protein is termed denatured. For small, monomeric proteins, the unfolding transition has been shown to be of a two-state nature; the protein exists in either a fully-folded (native) or a fully-unfolded (random coil) conformation. The unfolding of large and/or oligomeric proteins is more complex as each domain may unfold under differing conditions and the complete unfolding to a random coil conformation is unlikely (i.e., some residual structure will remain).

The area of an enzyme most sensitive to unfolding is likely to be the active site [Tsou, 1986]. To perform a desired reaction, specific residues must be in optimal alignment, as seen in the catalytic triad of serine proteases and the active site His residues and Asp of citrate synthase. In addition, the active site is normally present on a flexible region of the protein to allow substrate-entry and product-release. Thus, the overall design of an enzyme must compensate for the destabilisation conferred by such a hinge region. The consequences of this for a hyperthermostable enzyme are two-fold. First, the enzyme must maintain the native conformation at the growth temperature of the organism and must exhibit a similar degree of polypeptide flexibility and catalytic rate at the extreme temperature that an enzyme, from a mesophilic source, exhibits at the lower temperature. Second, even though the rate of diffusion of substrates and products in and out of the active site will be increased at the elevated temperature, the possible thermolability of substrates and co-factors may in fact necessitate further adaptation of the binding sites to increase the affinity of the enzyme for such molecules.

The precise nature of the interactions responsible for such a rigid (at ambient temperature) structure are not going to be easily pinpointed. Experiments involving the guanidine or urea-unfolding of small monomeric enzymes (M_r 12-15 kDa) have shown that the difference in free energy (ΔG) between the folded and the unfolded state is not likely to exceed 67 kJ/mol [Jaenicke, 1991]. It was then postulated that the amount of free energy change ($\Delta\Delta G$), required to maintain the native conformation at an elevated temperature would amount to the equivalent of just a few ion pairs, hydrogen bonds or hydrophobic patches. Thus, as every amino acid makes local and non-local contacts in a protein, every interaction has an effect on the overall stability of the protein, and a strained conformation at one location may be compensated for elsewhere in the molecule. It thereby follows, that the changes in sequence that enhance the stability of the folded conformation are likely to be subtle and distributed throughout the molecule. The assessment of the contribution of individual amino acids to the stability of the native state has been approached in two ways. The most extensive study, which will be described first, is by the site directed mutagenesis of small, monomeric proteins.

Most monomeric enzymes unfold reversibly, allowing the degree of unfolding at different concentrations of denaturant to be followed by biophysical methods such as CD or fluorescence spectroscopy. From a ratio of the fraction of the unfolded enzyme to the fraction folded, the free energy of folding (ΔG) under the applied conditions can be determined. The value of ΔG for the mutant enzyme can be compared with that of the wild-type enzyme under the same experimental conditions to assess the relative effect of the change on stability. Double mutant cycles are a variation of this procedure whereby a specific interaction between two residues can be assessed by mutating them separately and then together and assessing the unfolding characteristics of each [Kellis *et al.*, 1988; Serrano *et al.*, 1990]. A number of enzymes have been investigated extensively by these methods. These include ribonuclease from *Bacillus amyloliquefaciens* (barnase), staphylococcal nuclease, bacteriophage T4 lysozyme and *E. coli* ribonuclease HI (RNase HI). From these experiments, a number of putative stabilising features have been elucidated; these are discussed in detail below.

1.4.2 Protein stability predictions from site directed mutagenesis studies of small, monomeric enzymes

In a recent review, Fersht and Serrano [1993] derived a number of putative stabilising principles from the mutagenesis and urea or guanidine-unfolding equilibria studies. These studies have concentrated mainly on factors in α-helices that affect protein stability. In many cases, the mutations were sequence and context dependent and the ability of the protein to incorporate the change without causing dihedral angle strain elsewhere in the structure was of paramount importance.

Factors in α-helices that affect stability

Several factors affect α-helix stability including the electrostatic interactions between the side chains of residue *i* of an α-helix with residue *i+3*, hydrogen bonds between the backbone carbonyl oxygen of residue *i* and the amino group hydrogen of residue *i+4*, the α-helix dipole generated by such a hydrogen bonding pattern, hydrophobic and aromatic interactions and solvent accessibility of hydrophilic and hydrophobic residues.

Ala has been shown in both protein and model α-helices to have the highest α- helical propensity [Serrano *et al.*, 1992a]. Pro and Gly are most destabilising in an α-helix due to the loss of two hydrogen bonds and the greater freedom of bond rotation, respectively. Branched side chains may also destabilise α-helices as the number of allowed conformations in an α-helix is less than allowed in the random coil

conformation. Another factor to consider is the ability of the side chain to hydrogen bond to the main chain in an unfolded state which in turn, would contribute to stabilise this state. For this reason, Tyr may be more destabilising than Phe, Ser more than Thr and Asp more than Glu. The longer the side chain, the greater the entropic cost of hydrogen bonding to the main chain [Horovitz *et al.*, 1992; Serrano *et al.*, 1992b].

The context dependence of these studies is strongly highlighted in the assignment of preferred residues at the α-helix caps. An α-helix capping residue is defined as the first residue in an α-helix whose C_α is within the cylinder defined by the α-helix [reviewed by Hol, 1985]. These residues are thought to be important to stability as they counteract the α-helix dipole. Thus, mutation studies on barnase indicated that positively-charged residues, Arg and His were favoured at the carboxyl-cap whereas Asp, Ser and Thr were favoured at the amino-cap. Gly ranks highly for both caps and is the most likely residue to be found at the carboxyl-cap as it can adopt the appropriate dihedral angles to allow maximum solvation of the end of the helix [Serrano *et al.*, 1992b]. A similar series of mutations of T4 lysozyme illustrates how mutation studies are context dependent [Bell *et al.*, 1992]. Close examination of conflicting data for the preferred residues at the amino cap of an α-helix showed than a nearby Asp at N+2 (N being the amino-cap residue) was acting as a surrogate cap [Fersht and Serrano, 1993].

Other interactions in α-helices include surface and buried salt bridges. Double mutant cycles were used to analyse the strength of a salt bridge interaction on the surface of both T4 lysozyme and barnase [Dao-pin *et al.*, 1991; Šali *et al.*, 1991]. In both cases the actual contribution to stability was small, a property attributed to the entropic cost of fixing the two side chains such that an interaction can occur. Similar studies on buried salt bridges, which are rare in proteins, have shown that removing one partner of the interaction is highly destabilising, resulting in large rearrangements of the protein structure.

Internal factors that affect stability

The internal hydrophobic core of an enzyme has been studied in *E. coli* RNase HI, barnase, staphylococcal nuclease and T4 lysozyme [Ishikawa *et al.*, 1993; Kellis *et al.*, 1988; Shortle *et al.*, 1990; Eriksson *et al.*, 1992]. In the former two systems, attempts were made to 'fill' internal cavities by substituting a smaller aliphatic residue with a bulkier one. The opposite approach was adopted in the latter two systems; substitutions were made that effectively truncated the residue in the native structure, thus creating a cavity.

Cavity filling may cause dihedral angle strain or more extreme changes to the overall structure. Attempts to stabilise T4 lysozyme by a Leu to Phe and an Ala to Val mutation did not produce significant stabilisation [Eriksson *et al.*, 1992]. However, stabilisation was achieved in *E. coli* RNase HI by a Val to Ile and a Val to Leu substitution [Ishikawa *et al.*, 1993]. This suggested that in the latter system, the mutation was incorporated into the enzyme structure without dihedral angle strain and that the enhanced van der Waals interactions increased the stability of the folded conformation. The effective truncation of an amino acid side chain is destabilising to an enzyme. This is expected as the enzyme has evolved to maximise the favourable van der Waals contacts which are being disrupted. On assessing a total of 83 mutants of staphylococcal nuclease, Shortle *et al.* [1990] demonstrated that the relative stability of a particular cavity-creating mutant was dependent on the ability of the residues lining the cavity to collapse into it and minimise its size.

The internal hydrophobic core is thought to be of fundamental importance to protein folding. On exposure to solvent, the ordering of water molecules in a cage-like structure around hydrophobic side chains is of such entropic cost to the system that the internalisation of such regions is thought to be the dominant folding force [Dill, 1990]. Another putative determinant of the folding pathway is the ubiquitous hydrogen bond, the role of which in α-helix stability was discussed earlier. The contribution of

hydrogen bonding to the stability of the folded state was studied in *E. coli* RNase TI, a molecule with 86 intramolecular hydrogen bonds [Shirley *et al.*, 1992]. The removal of one partner of a hydrogen bonding interaction is highly destabilising, especially if the bond occurs in a hydrophobic environment and the strength of the electrostatic interaction is thereby enhanced. By investigating the removal of the hydrogen bonding capacity of 12 different residues of RNase TI, Shirley and co-workers [1992] suggested that hydrogen bonding plays a comparable role to the hydrophobic interaction in protein folding.

Interactions between aromatic residues

The final categories of interaction that have been studied by this approach are those between aromatic residues and between aromatic residues and His residues. Both of these interactions have been studied in barnase [Serrano *et al.*, 1991]. Two aromatic residues in close proximity are most stable in a tilted T arrangement whereby the π electrons and the partially negatively-charged carbon atoms of the face of one aromatic moiety can interact with the partially positively-charged hydrogen atoms on the edge of the adjacent aromatic side chain [Singh and Thornton, 1985; Burley and Petsko, 1985]. The negatively-charged face of an aromatic ring also serves to stabilise a nearby His residue in a protonated form [Loewenthal *et al.*, 1992]. This interaction may have a significant role in the stabilisation of a protonated His required in a catalytic mechanism such as a charge relay system.

Overall, the major force driving the unfolding of a protein is thought to be the favoured conformational entropy of the unfolded state, as there is minimal restriction of rotation around covalent bonds. This is overcome in the native protein by the stabilising interactions of hydrogen bonding, van der Waals forces and electrostatics, in addition to the overall entropic gain of removing the hydrophobic side chains from

exposure to the solvent and the subsequent release of the ordered water surrounding these regions [Dill, 1990]. These detailed and extensive studies have highlighted the contribution of different residue types at internal and external positions of a protein to the stability of the native state. However, to complement this work, it is essential to investigate the mechanisms that are actually used in nature to increase the stability of the folded state such that it is maintained under extreme environmental conditions. To do this, the crystal structures of enzymes from extremophile sources must be pursued, compared to the non-thermostable equivalents and potentially stabilising features tested by site direct mutagenesis. Thus, the second approach is the investigation of naturally-occurring hyperthermostable enzymes.

1.4.3 Protein stability predictions by comparative enzymology

The study of protein thermostability by the investigation of naturally-occurring hyperthermostable enzymes is less advanced than the above study for two reasons. First, hyperthermophilic organisms have only recently been isolated and cultured on a large scale. Second, the difficulties in culturing these organisms (e.g. temperature, anaerobicity and fermenter-deterioration from H_2S expulsion) has necessitated the cloning and over-expression of the genes of interest in hosts such as *E. coli*, thus delaying a detailed structure-function analysis. However, the progress made in using this approach is described below.

Predictions from amino acid sequence comparisons

In the absence of any crystal structures of thermostable proteins, early comparative analyses focused solely on trends of 'cold' to 'hot' amino acid substitutions and then attempted to rationalise the preference in physical terms. The most extensive study was undertaken by Menéndez-Arias and Argos [1989] who analysed 65 enzyme sequences from 6 protein families that contained at least one thermostable counterpart. It must be noted that the highest optimal growth temperature in this study was just 80°C and all species have been shown by phylogenetic analysis to be bacterial or eukaryal.

The ten most frequent 'cold' to 'hot' replacements were Lys-Arg, Ser-Ala, Gly-Ala, Ser-Thr, Ile-Val, Lys-Ala, Thr-Ala, Lys-Glu, Glu-Arg and Asp-Arg. The substitutions to Ala are thought to confer stability because this residue exhibits the highest α-helical propensity. Substitutions involving the addition of methyl groups (e.g. Ser-Thr) are believed to be stabilising by increasing the hydrophobicity of the side chain and by bestowing a greater entropic cost of hydrogen bonding to the main chain. For this reason, the substitution of Asp to Glu is also thought to be thermostabilising although it does not feature in the top ten amino acid exchanges. A solvation effect is believed to be the underlying cause of the Lys-Arg trend as controlled guanidination of solvent-exposed Lys residues has been demonstrated to increase the thermostability of some proteins [Qaw and Brewer, 1986]. As mentioned in Section 1.4.2, Gly is discriminated against in thermostable proteins as the number of conformations it can adopt greatly favours the unfolded state. Other sequence changes have been predicted following the analysis of irreversible inactivation of proteins at high temperature. The compositional changes involve a general decrease in Cys residues which are susceptible to oxidation [Zale and Klibanov, 1986], Asn and Gln which can be easily deamidated [Geiger and Clarke, 1987; Stephenson and Clarke, 1989] and Asp residues which can cause cleavage of the polypeptide chain in weakly acidic conditions [Ahern and Klibanov, 1985; Zale and Klibanov, 1986].

The adherence to the 'traffic rules' of Menéndez-Arias and Argos [1989] in thermostable proteins has not been absolute. For example, GAPDH has been studied from *Bacillus stearothermophilus*, *Thermotoga maritima* and *P. woesei* (T_{opt} = 52.5°C, 80°C and 100°C, respectively) [Davies *et al.*, 1991; Schultes *et al.*, 1990; Zwickl *et al.*, 1990]. The *P. furiosus* and *B. stearothermophilus* enzymes showed a marked decrease in Cys residues, possessing just the one catalytic Cys whereas the *Thermotoga* enzyme contained 3 Cys residues. In addition, whereas the *Thermotoga* citrate synthase obeyed the Lys-Arg, Leu-Ile and Ser-Thr traffic rules but had average numbers of Gly, Ala and Val, the *P. woesei* enzyme had indifferent Arg, Lys, Ser and Thr values but showed a preference for Ala, Ile and aromatic residues and a discrimination against Gly. Similarly, glutamate dehydrogenase (GDH) of *P. furiosus* exhibited a decrease in Gly and Cys and an increase in Ile and Asp but no significant changes in Arg, Lys, Pro or Gly [Eggen *et al.*, 1993]. Hence, hyperthermostabilisation of proteins may involve different amino acid trends than stabilisation to moderate temperatures.

As the following section will demonstrate, the lack of adherence to these rules can often be explained by the use of the 'unfavoured' substitution in the formation of a new salt bridge or hydrogen bond. Thus, in recent years, the traffic rules have been considered to be of limited value apart from constituting a starting point for a comparative investigation. Upon the accumulation of more sequence data from archaeal hyperthermophiles, a revised set of rules may be formulated which are more specific to this phylogenetic group.

Predictions from crystal structure comparisons

Four comparative studies between thermostable and non-thermostable protein structures have extended to the level of crystallographic analysis. Only one of these, the 7 kDa redox co-factor, rubredoxin, has been elucidated from a hyperthermophilic Archaeon [Blake *et al.*, 1991]. However, despite the variety in the phylogenetic position of the selected enzymes, a number of correlations have been observed between three of the studies.

The least informative structure is actually the most thermostable. Perhaps a function of its small size, rubredoxin from *P. furiosus* exhibited just one obvious difference that may account for its thermostability. The anchoring of the amino-terminus to another region of secondary structure by hydrogen bonding was highlighted in this co-factor. This feature was also seen in two other structural comparisons, the comparison of the neutral protease of *Bacillus cereus* with thermolysin of *Bacillus thermoproteolyticus* [Pauptit *et al.*, 1988] and the comparison of 3-phosphoglycerate kinase (PGK) from yeast with that of *Bacillus stearothermophilus* [Davies *et al.*, 1993]. The former exhibited an extra salt bridge fixing the carboxyl-terminus whereas both the amino- and carboxyl-termini of PGK were linked to secondary structure elements by such interactions.

The comparison of neutral protease with thermolysin also explained the adherence (or lack of adherence) to the 'traffic rules' described in the previous section. Thermolysin exhibited the substitutions of Lys by Arg at solvent-exposed positions, but of the 9 Ser-Ala substitutions, 7 were in the opposite direction to that proposed by Menéndez-Arias and Argos [1989]. An analysis of the position of these residues in the

crystal structure showed that the two substitutions in the Ser-Ala direction were in regions of α-helix, whereas of the 7 Ala-Ser replacements, 6 resulted in the formation of new hydrogen bonds and 1 was in a solvent-exposed position. Thus in context, these residues were more likely to contribute to the stability of the folded conformation, despite not obeying the general trends.

The fourth crystal structure comparison was that of the citrate synthase of pig with the moderately thermophilic Archaeon, *T. acidophilum* [Russell *et al.*, 1994] (Section 1.1.4). This comparison correlated with the PGK and thermolysin studies in that the domains encompassing the active site were orientated towards a more closed conformation in the thermostable enzyme. This infers again, that enzyme compactness is central to thermostability. Other features of the *T. acidophilum* enzyme are described in Section 1.1.4.

These studies indicate that general trends in amino acid substitutions and results of mutagenesis studies are a useful starting point for a comparative study but as each residue can make local and non-local contacts, each substitution is context dependent, necessitating the determination of the crystal structure. Hence, the nature of the work reported in this thesis is to pursue this situation with the hyperthermostable citrate synthase of *P. furiosus*.

1.5 THE AIMS OF THIS PROJECT

To remain functional, hyperthermostable proteins must withstand unfolding to a non-native conformation and avoid irreversible inactivation at extreme temperatures. These are not mutually exclusive as the exposure of a susceptible linkage to water and the achievement of dihedral angles optimum for a deteriorative reaction are a direct function of the stability of the folded state. As mentioned in Section 1.4.1, the hyperthermostable enzyme must exhibit a similar degree of flexibility at the extreme temperature that the enzyme from a mesophilic source exhibits at the lower temperature. This is also important for the cellular life of an enzyme. If the enzyme is too rigid at the operational temperature, the organism may not be able to conserve resources by degrading non-essential enzymes. This may be especially true for the *P. furiosus* citrate synthase as it is likely to have a predominantly biosynthetic role.

Thus, the protein honed for thermostability is likely to be rigid at room temperature, a situation conferred intrinsically by the primary sequence. Also, to lower the possibility of irreversible inactivation, the hyperthermostable enzyme may avoid potential thermolabile residues such as Cys, Asp and Asn. The accurate assessment of protein thermostability requires the determination of the crystal structure of the native enzymes. However, due to difficulties in culturing these organisms, this invariably requires the cloning, sequencing and over-expression in *E. coli* of the relevant genes.

Hence, the aim of this study is to purify citrate synthase from *P. furiosus* and to use the amino-terminal sequence to design an oligonucleotide probe to clone the citrate synthase gene. The aim is then to express the gene and produce adequate quantities of functional, hyperthermostable enzyme for thermostability and preliminary crystallographic analyses. The sequence data obtained from the cloned gene can then be analysed for putative thermostabilising features and used to obtain a protein-derived phylogenetic analysis of citrate synthase sequences.

CHAPTER 2

GENERAL METHODS

This chapter details the general techniques used in the experiments described in this thesis. Methods that are specific to a particular chapter are described within that chapter.

2.1 ENZYMOLOGY AND PROTEIN METHODS

2.1.1 Materials

CoA was supplied by Calbiochem-Novobiochem, Nottingham, UK. Oxaloacetate, 5,5'-di-thiobis-(2-nitrobenzoic acid) (DTNB), N,N,N',N'-tetramethylethylene diamine (TEMED), Coomassie Brilliant Blue R and Coomassie Brilliant Blue G-250 were supplied by Sigma Chemical Co. Ltd., Poole, UK. Protogel™ was supplied by National Diagnostics Ltd., New Jersey, USA and protein molecular weight standards from Bio-rad Laboratories, Herts., UK. Polyvinylidene difluoride (PVDF) membrane, Immobilon™ P, was obtained from Millipore, Watford, UK and 3MM paper from Whatman Scientific Ltd., Maidstone, UK. Bovine serum albumin, (BSA) was obtained from Pharmacia, St. Albans, UK. Standard chemicals and solvents of the highest quality available were supplied by either Sigma, Fisons, Loughborough, UK., or BDH Ltd., Poole, UK.

2.1.2 Spectrophotometric assay for citrate synthase

Citrate synthase was assayed at 412 nm and 55°C by the detection of the reaction product, CoA, using the chromogenic, thiol-specific reagent, DTNB [Srere *et al.* 1963]. Unless stated otherwise, the assay was performed in plastic cuvettes, in a 1 ml volume of 2x TE buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA) containing acetyl-CoA to a

concentration of 0.15 mM, DTNB to 0.1 mM and oxaloacetate to 0.2 mM. Acetyl-CoA was prepared by adding, to a 10 mg/ml pre-chilled solution of CoA, $KHCO_3$ to a concentration of 0.17 M (from a 1 M stock solution) and acetic anhydride to 11 mM. Non-enzymic generation of CoA was measured by omitting oxaloacetate from the assay.

2.1.3 Determination of protein concentration

Protein concentration was determined by the method of Bradford [1976]. A 0.9 ml volume of filtered Bradford reagent (0.01 % (w/v) Coomassie G-250, 4.8 % (v/v) EtOH, 8.5 % (v/v) H_3PO_4) was incubated with a 0.1 ml volume of protein sample at room temperature for 15 min. The absorbance of the sample at 595 nm was then measured against a reference cuvette containing a 0.9 ml volume of Bradford reagent and a 0.1 ml volume of water. Protein concentration was determined by comparing the sample absorbances to a standard curve prepared, as above, using concentrations of BSA from 0-30 μg/ml.

2.1.4 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were routinely analysed by the SDS-PAGE method of Laemmli [1970]. A 20 ml volume of resolving gel solution (10 % (w/v) acrylamide (Protogel), 0.375 M Tris-HCl pH 8.8, 0.1 % (w/v) SDS) was polymerised by the addition of 5 mg ammonium persulphate (from a 10 % (w/v) solution) and a 20 μl volume of TEMED. The gel solution was then poured into a gel mould, overlaid with *n*-butanol and left at room temperature until visibly set. The *n*-butanol was removed and the upper edge of the resolving gel washed with distilled water before being overlaid with a stacking gel. A 10 ml volume of stacking gel solution (5 % (w/v) acrylamide (Protogel), 0.13 M Tris-HCl pH 6.5, 0.1 % (w/v) SDS) was polymerised with 3 mg ammonium persulphate (from a 10 % (w/v) solution) and a 10 μl volume of TEMED. A well-forming comb was positioned in the stacking gel which was left to set at room temperature. The comb was then removed and

the gel placed in an electrophoresis tank. The upper and lower tanks were filled with electrophoresis buffer (0.025 M Tris-HCl pH 8.8, 0.096 M Gly, 0.1 % (w/v) SDS) and bubbles removed from the lower edge of the gel. Unpolymerised acrylamide was washed from the sample wells with electrophoresis buffer before loading the protein samples.

Protein samples were prepared for electrophoresis by adding an equal volume of sample buffer (0.18 M Tris-HCl pH 6.5, 2 % (w/v) SDS, 10 % (w/v) glycerol, 2.6 % (w/v) β-mercaptoethanol, 0.005 % (w/v) bromophenol blue) and boiling for 2 min. Very dilute protein samples were precipitated by adding trichloroacetic acid (TCA) to 10 % (w/v), incubating on ice for 30 min, and centrifuging for 10 min at 13,000 x *g*. The protein pellet was washed with acetone, dried under vacuum, resuspended in sample buffer and boiled for 2 min. Samples were electrophoresed at a constant current of 10 mA through the stacking gel and 20 mA through the resolving gel until the bromophenol blue dye front was 0.5 cm from the lower edge of the gel. Following electrophoresis, gels were stained for 30 min with 0.5 % (w/v) Coomassie Brilliant Blue R in 9:2:9 (v:v:v) methanol:acetic acid:water and destained with 2:1:7 (v:v:v) methanol:acetic acid:water.

2.1.5 Electrophoretic transfer of proteins from polyacrylamide gels to PVDF membranes

The protein samples were electrophoresed through a 10 % (w/v) polyacrylamide gel (Section 2.1.4) which, without staining, was prepared for semi-dry protein blotting by incubating in transfer buffer (50 mM Tris, 50 mM Gly, 20 % (v/v) methanol) for 20 min. The PVDF membrane (Immobilon P), cut to the same size as the gel, was submerged in methanol for 3 sec, water for 1-2 min and transfer buffer for 3 min. Fourteen squares of Whatman 3MM chromatography paper, cut to the same size as the gel, were soaked in transfer buffer before assembling the semi-dry electroblotting apparatus. The electroblot was assembled in a Novablot apparatus (Pharmacia) in the following order: positive electrode, 7 filters, 2 PVDF membranes, gel, 7 filters, negative electrode. Transfer of protein bands to the membrane was achieved by applying a constant current of 0.8 mA/cm^2 membrane for 2 h. The membrane was then stained with 0.025 % (w/v) Coomassie

Brilliant Blue R in 10 % (v/v) acetic acid, 50 % (v/v) methanol, and destained with a solution containing 10 % (v/v) acetic acid and 50 % (v/v) methanol. Protein bands that required amino-terminal sequencing were cut out and air dried.

2.1.6 Protein microsequencing

Microsequencing was performed by on an Applied Biosystems 470A gas-phase sequence coupled to an Applied Biosystems 120 phenylthiohydantoin analyser (Mrs J. Young Zeneca Pharmaceuticals, Macclesfield).

2.2 MOLECULAR BIOLOGY METHODS

2.2.1 Enzymes, reagents and other materials

Restriction enzymes, polynucleotide kinase, DNA polymerase I, Sepharose® CL-6B and Sephadex® G-50 were supplied by Pharmacia, *Taq* DNA polymerase by Boehringer-Mannheim, Mannheim, Germany, and T4 DNA ligase by Stratagene, La Jolla, USA. $Vent_R$ DNA polymerase was obtained from New England Biolabs, Beverly, USA. Sequenase® 2.0 was supplied in a DNA sequencing kit from United States Biochemical Corporation, Ohio, USA. Sequagel™ reagents were obtained from National Diagnostics, Atlanta, USA. Agarose, ethidium bromide, ampicillin, glass wool and Sigmacote™ were obtained from Sigma. *Hin*dIII-digested and *Pst*I-digested λ DNA, isopropyl-β-D-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) were obtained from Northumbria Biologicals Ltd., Cramlington, UK. Constituents of *E. coli* culture media were supplied by Difco Laboratories, Michigan, USA. Genescreen™ membrane, [γ^{32}P] ATP and NENsorb™ DNA purification columns were supplied by Dupont-NEN Research products, Stevenage, UK. [α^{35}S] ATP was obtained from ICN Bio-medicals Ltd., High Wycombe, UK, and X-ray film from the Fuji Film Company, Japan. Geneclean® DNA

purification kits were obtained from Bio 101, La Jolla, USA. All reagents specific to non-radioactive labelling and detection of oligonucleotides using digoxygenin were supplied by Boehringer-Mannheim. Standard chemicals and solvents, of the highest quality available were obtained from Fisons, BDH, Sigma or Fluka, Poole, UK.

2.2.2 Bacterial strains, culture conditions and plasmids

Escherichia coli strain XL1 Blue (*recA1*, *endA1*, *gyrA46*, *thi*, *hsdR17*, *supE44*, *relA1*, *lac* F'[*proAB*$^+$ *lacI*^{q} *lacZ*Δ *M15 Tn10*(*tet*^{r})]) was obtained from Stratagene. *E. coli* strain JM105 (*endA*, *sbcB15*, *hsdR4*, *rpsL*, *thi* Δ(*lac-proAB*) F'[*traD36 proAB*$^+$ *lacI*^{q} *lacZΔM15*]) was used for expression studies.

E. coli strains were cultured at 37°C in 2xYT liquid medium (16 g/l bactotryptone, 10 g/l yeast extract, 5 g/l NaCl) or, on LB-agar solid medium (10 g/l bactotryptone, 5 g/l NaCl, 5 g/l yeast extract, 15 g/l agar). For the propagation of *E. coli* strains carrying plasmids conferring ampicillin resistance, media was supplemented with ampicillin to a concentration of 100 μg/ml. Long term storage of *E. coli* cultures was achieved by incubating cells of an overnight culture in 50 % (v/v) glycerol at -20°C. The plasmids used in this study, pUC18, pUC18-*Sma*I/BAP and pKK223-3, were obtained from Pharmacia.

2.2.3 Quantitation of DNA

The concentration of a solution of DNA was determined by measuring the absorbance at 260 nm [Sambrook *et al.*, 1989]. For double-stranded DNA, a 50 μg/ml solution has an absorbance of 1 unit at 260 nm. Purity of a DNA sample was assessed by calculating the A_{260}/A_{280} ratio of a sample; a ratio of 1.8 or more indicates that the sample is pure.

2.2.4 Precipitation of DNA with ethanol

Precipitation with ethanol was used either to concentrate a DNA sample or to effect a buffer-exchange. DNA was precipitated by adding sodium acetate to a concentration of 0.3 M from a 3 M solution at pH 5.2, and three volumes of pre-chilled absolute ethanol. The sample was incubated at -20°C for 30 min and the DNA pelleted by centrifugation at 13,000 x *g* for 10 min. The DNA pellet was then washed with 70 % (v/v) ethanol, pelleted, dried under vacuum and resuspended in the desired volume of buffer.

2.2.5 Digestion of DNA with restriction endonucleases

DNA was digested with the desired restriction endonucleases as detailed by the manufacturer. The enzyme and DNA were incubated at 37°C in 1x or 2x One-Phor-All buffer (supplied by the manufacturer). When digesting DNA with two enzymes, the enzymes could be used simultaneously if the buffer requirements were the same. However, if different buffer conditions were required, the DNA was digested first with the enzyme requiring the lower buffer concentration. This enzyme was then inactivated by incubating for 30 min at the inactivation temperature (determined by the manufacturer), cooled to room temperature and the buffer concentration adjusted before adding the second enzyme.

2.2.6 Agarose gel electrophoresis

DNA fragments were routinely analysed by horizontal, agarose gel electrophoresis using 0.7 % (w/v) agarose gels which provide efficient separation of linear, duplex DNA of 0.8-10.0 kb [described by Sambrook *et al.*, 1989]. Agarose was dissolved in the desired amount of 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA) by heating. When cooled to approximately 45°C, ethidium bromide was added to a concentration of 0.5 μg/ml. The warm agarose was then poured into a perspex gel mould, a well-forming comb clamped into place, and the gel left at room temperature to set. Once set, the comb was removed

and the gel placed in a gel tank and covered with 1x TAE buffer (40 mM Tris-acetate, 2 mM EDTA). Samples, in 1x loading buffer (0.04 % (w/v) bromophenol blue, 0.04 % (w/v) xylene cyanol, 5 % (v/v) glycerol), were transferred to the sample wells and electrophoresed at a constant voltage of 60 V until the xylene cyanol dye front was 1 cm from the lower edge of the gel. The size-separated DNA was then visualised under UV transillumination.

2.2.7 Purification of DNA from agarose gels

Two methods were used to extract DNA from agarose gels: the freeze-squeeze method which is based on the method of Tautz and Renz [1983] and the Geneclean method. To perform the freeze-squeeze method, the excised fragment was placed into a 0.5 ml microfuge tube which had been pierced, and plugged with silane-treated glass wool. The microfuge tube and contents were then submerged in liquid nitrogen for 1 min, placed inside a 1.5 ml microfuge tube, and centrifuged at 13,000 x *g* for 10 min. By this, a solution of the DNA collected in the outer 1.5 ml microfuge tube while the agarose was retained by the glass wool. The DNA was then purified by passage though a NENsorb column as described in the manufacturer's instructions.

The Geneclean method relies on the affinity of DNA for a silica based matrix, 'glassmilk'. The DNA was purified from agarose and other contaminants as described in the manufacturer's instructions.

2.2.8 Southern blotting of DNA from agarose gels onto nylon membranes

The DNA sample was size-separated by agarose gel electrophoresis, as described in Section 2.2.6. Following electrophoresis, the gel was photographed under transillumination and prepared for Southern blotting [Southern, 1975]. The gel was incubated for 15 min in a solution of 0.25 M HCl and then for 30 min in a solution of 0.2 M NaOH containing 0.6 M NaCl. It was then equilibrated in three changes of 10x SSC buffer (1.5 M NaCl,

0.15 M trisodium citrate), incubating each for 20 min. The Southern blot was then prepared as follows. An inverted gel casting tray was used as a raised platform within a shallow glass dish, filled to a depth of 2.5 cm with 10x SSC buffer. Two strips of Whatman 3MM paper were placed over the platform, to serve as a wick, onto which the gel, membrane and two filter papers were placed. Both the membrane and the filters had been previously cut to the same size as the gel and pre-wetted with 2x SSC buffer. Bubbles were removed at each stage to promote even transfer of DNA. Capillary transfer of buffer through the gel, membrane and filters was then effected by placing a 5-8 cm stack of paper towels on top of the filters and securing the stack with a 500 g weight. The apparatus was left overnight and the transferred DNA fixed to the membrane by baking at 80°C for 1-2 h.

2.2.9 Synthesis and purification of oligonucleotide probes

Oligonucleotides were synthesised using an Applied Biosystems 381A DNA synthesiser (A. Wolstenholme, University of Bath). The solid phase support, containing the synthesised oligonucleotide, was incubated overnight at 55°C in a 1 ml volume of ammonium hydroxide. Following the removal of the solid phase support by centrifugation at 13,000 x *g* for 2 min, the ammonium hydroxide was neutralised with an equivalent amount of glacial acetic acid and the DNA precipitated as described in Section 2.2.4. The concentration of the oligonucleotide was determined as described in Section 2.2.3.

2.2.10 5' End labelling of the oligonucleotide with [γ^{32}P] ATP

A 20 pmole quantity of oligonucleotide was incubated with 70 μC_i [γ^{32}P] ATP and 14 U T4 polynucleotide kinase in a 40 µl volume of 1x kinase buffer (supplied by manufacturer) at 37°C for 30 min. The labelled oligonucleotide was then separated from unincorporated [γ^{32}P] ATP by passage through a 0.5 ml Sephadex G-50 column. This was prepared in a microfuge tube, of volume 0.5 ml, that had been pierced and plugged with silane-treated glass wool. The Sephadex G-50 storage buffer was removed by centrifuging

at 3,000 x *g* for 2 min and the matrix washed with 1x TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). After centrifuging the column to dryness, the labelling mixture was applied and the labelled oligonucleotide separated from unincorporated [γ^{32}P] ATP by a final centrifugation into a microfuge tube of volume 1.5 ml.

2.2.11 3' Tailing of the oligonucleotide with digoxygenin-11-dUTP/ATP

As an alternative to radioactive labelling, the probe was tailed with digoxygenin (DIG)-dUTP/dATP. A 100 pmole quantity of probe was labelled according to the manufacturer's instructions.

2.2.12 Hybridisation and detection of [γ^{32}P]-labelled oligonucleotide probes

The DNA sample was size-separated by agarose gel electrophoresis and transferred by the method of Southern [1975] onto a nylon membrane (Section 2.2.8). The membrane was incubated for 6 h at the assigned temperature, in prehybridisation solution (6x SSC buffer, 0.5 % (w/v) SDS, 100 µg/ml denatured salmon sperm, 5x Denhardt's reagent (1g/l Ficoll, 1g/l polyvinylpyrrolidone, 1g/l BSA)). This solution was then replaced with fresh prehybridisation solution containing labelled oligonucleotide at a concentration of 1-2 pmoles/ml and the membrane incubated overnight at the assigned temperature. Following hybridisation, the membrane was washed twice, for 5 min at the assigned temperature, with 2x SSC buffer containing 0.1 % (w/v) SDS, and twice, for 20 min, with 0.1x SSC buffer containing 0.1 % (w/v) SDS. The membrane was then wrapped in clingfilm and exposed overnight at -70°C to X-ray film (Fuji RX100). The X-ray film was developed using an Amersham Hyperprocessor.

All incubations were performed at a temperature specific to a particular oligonucleotide. The temperature at which an oligonucleotide specifically bound to the target DNA was estimated, by calculating the approximate melting temperature for the oligonucleotide, and then refined experimentally. The melting temperature of an oligonucleotide was estimated by assigning every possible A-T base pair between the oligonucleotide and the target DNA a value of 2°C, and every G-C base pair a value of 4°C [Wallace *et al.*, 1979].

2.2.13 Hybridisation and detection of digoxygenin-labelled oligonucleotide probes

The DNA sample was size-separated by agarose gel electrophoresis and transferred by the method of Southern [1975] onto a nylon membrane (Section 2.2.12). The membrane was incubated in a prehybridisation solution of 5x SSC buffer containing 1 % (w/v) casein blocking reagent, 0.1 % (w/v) N-lauroylsarkosine, 0.2 % (w/v) SDS and 5 μg/ml polydeoxyadenosine. After 2 h, this was replaced with fresh prehybridisation solution containing labelled oligonucleotide at a concentration of 1-2 pmoles/ml and incubated for a further 2 h. Both prehybridisation and hybridisation steps were performed at a temperature specific to a particular oligonucleotide, as detailed in Section 2.2.12. Following hybridisation, the membrane was washed twice, for 5 min at the assigned temperature, with 2x SSC buffer containing 0.1 % (w/v) SDS, and twice, for 20 min, with 0.1x SSC buffer containing 0.1 % (w/v) SDS.

Hybridised oligonucleotide was detected as described in the manufacturer's instructions. Detection was achieved using an alkaline phosphatase-conjugated, anti-DIG fragment antibody. The chemiluminescent reagent, lumigen PPD™ was used as a substrate for the conjugated enzyme, and the treated membrane exposed to X-ray film for 30 min-2 h. The X-ray film was then developed using an Amersham hyperprocessor.

2.2.14 Preparation and transformation of competent *E. coli* (XL1 Blue and JM105)

Competent cells were prepared by the method of Hanahan [1985]. A saturated culture of *E. coli* (XL1 Blue or JM105, as specified) was prepared by inoculating a 2 ml volume of 2xYT media with 10 µl of a cell glycerol stock and incubating overnight at 37°C. A 100 ml volume of 2xYT media was then inoculated with 0.1 ml of the saturated culture and incubated at 37°C until the absorbance of the culture at 550 nm was 0.4 units. The cells were then centrifuged at 3,500 x *g* for 10 min and the cell pellet resuspended in a 40 ml volume of pre-chilled RFI solution (100 mM $RbCl_2$, 50 mM $MnCl_2$, 30 mM $K^+CH_3COO^-$, 10 mM $CaCl_2$, 15 % (v/v) glycerol, pH 5.8 (acetic acid)). After incubating on ice for 5 min, the cells were pelleted and resuspended in an 8 ml volume of pre-chilled RFII solution (10 mM MOPS, 10 mM $RbCl_2$, 75 mM $CaCl_2$, 15 % (v/v) glycerol pH 6.8 (NaOH)). The cells were then incubated overnight at 0°C and stored at -70°C after flash-freezing in liquid nitrogen in aliquots of 200 µl.

To transform the cells, a 200 µl aliquot was thawed on ice and incubated with plasmid DNA on ice for 30 min. The cells were then subjected to a heat shock of 42°C for 90 sec and immediately returned to ice for 2 min. A 0.8 ml volume of 2xYT media was added to the cells which were then incubated at 37°C for 1 h to allow phenotypic expression of the β-lactamase gene. Aliquots of the cells were then spread on LB-agar plates, supplemented with ampicillin to a concentration of 100 µg/ml.

2.2.15 Small scale preparation of plasmid DNA - 'Miniprep'

E. coli (XL1 Blue or JM105, as specified), containing the plasmid of interest, were cultured overnight, at 37°C, in a 3 ml volume of 2xYT media containing ampicillin. Plasmid was prepared by a method based on the alkaline lysis method of Sambrook *et al.*, [1989]. A 1 ml volume of the overnight culture was centrifuged at 13,000 x *g* for 5 min and the cells resuspended in a 150 µl volume of pre-chilled lysis buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl pH 8.0). A 300 µl volume of 0.2 M NaOH, 1 % (w/v)

SDS was added and the cells left on ice for 10 min. A 200 µl volume of 3 M potassium acetate pH 4.5 was added to the lysed cells and, after incubating for 20 min on ice, the precipitated material removed by centrifugation at 13,000 x *g* for 10 min. The supernatant was then extracted twice with an equal volume of 25:24:1 (v:v:v) phenol:chloroform:isoamylalcohol and the DNA precipitated with ethanol (Section 2.2.4). The DNA was then resuspended in a 50 µl volume of 1x TE buffer pH 8.0 and treated with 10 µg RNase A from a 10 mg/ml stock solution.

2.2.16 Large scale preparation of plasmid DNA - 'Maxiprep'

E. coli cells containing the plasmid of interest were cultured overnight, at 37°C, in a 100 ml volume of 2xYT medium, supplemented with 100 µg ampicillin/ml. Plasmid was then prepared by a method based on that described by Ausubel *et al.*, [1987]. The cells were centrifuged at 5,000 x *g* for 20 min and the cell pellet resuspended in a 5 ml volume of pre-chilled lysis buffer (25 mM Tris-HCl pH 7.5, 50 mM glucose, 1 mM EDTA). A 10 ml volume of 0.2 M NaOH containing 1 % (w/v) SDS was added and the cells gently agitated to effect cell lysis. Sodium acetate was added to the lysed cells to a concentration of 1 M (from a 3 M solution at pH 5.2) and the preparation incubated for 30 min on ice. The precipitated material was removed by centrifugation at 5,000 x *g* for 10 min. The cell supernatant was then added to a 25 ml volume of pre-chilled isopropanol and the preparation centrifuged again, as above. The pellet was dissolved in a 2 ml volume of 1x TE buffer pH 8.0 and a 2.5 ml volume of pre-chilled 5 M LiCl added. After incubating on ice for 5 min, the preparation was centrifuged at 5,000 x *g* for 10 min, the supernatant removed, and the DNA precipitated with absolute ethanol as described in Section 2.2.4.

The DNA pellet was air-dried and resuspended in a 0.5 ml volume of 1x TE buffer pH 8.0 containing 0.24 mg RNase A from a 10 mg/ml solution. After incubating at room temperature for 30 min, the DNA was precipitated by the addition of an equal volume of 1.6 M NaCl, 13 % (w/v) PEG 6000 and pelleted by centrifugation at 13,000 x *g* for 5 min. The pellet was resuspended in a 0.5 ml volume of 1x TE buffer pH 8.0 and extracted twice

with an equal volume of 25:24:1 (v:v:v) phenol:chloroform:isoamylalcohol. The plasmid DNA was then precipitated (Section 2.2.4), resuspended in 100 µl TE buffer pH 8.0, and quantified as described in Section 2.2.3.

2.2.17 Sequencing of DNA

The "maxiprep" method, described in Section 2.2.16 was used to prepare pure plasmid DNA. The insert was sequenced by the dideoxy-sequencing method of Sanger *et al.* [1977] using a Sequenase 2.0 kit. Prior to performing the sequencing reactions, it was necessary to denature the double-stranded plasmid. A 20 µg quantity of plasmid was incubated for 15 min at 37°C in a 40 µl volume of 0.2 M NaOH containing 0.2 mM EDTA. The denatured plasmid was then purified by passage through a column of Sepharose CL-6B of volume 0.5 ml. This was prepared in the same manner as the Sephadex G-50 column, described in Section 2.2.10. The denatured plasmid was applied to the column and separated from the NaOH and EDTA ions by a final centrifugation into a microfuge tube of volume 1.5 ml.

A 2 pmole quantity of the denatured plasmid was annealed to 2 pmoles primer (a single-stranded oligonucleotide of 17-25 bases in length, complementary to the target DNA) in 1x reaction buffer (supplied by the manufacturer). Annealing was achieved by heating to 65°C for 2 min and cooling to 30°C over 30 min. The extension and termination reactions were then performed as described in the manufacturer's instructions. The sequencing reactions were then analysed by electrophoresis through a polyacrylamide-urea gel prepared using Sequagel reagents. A Bio-Rad sequencing apparatus was prepared by cleaning both glass plates with ethanol and applying a 0.5 ml volume of Sigmacote to the back plate. The apparatus was then assembled as detailed by the manufacturer. Sequagel concentrate, diluent and buffer were mixed and polymerisation initiated with ammonium persulphate and TEMED as described in the manufacturer's instructions. A 21 x 50 x 0.4 cm sequencing gel was then poured and the well-forming comb placed in position.

Once set, the sequencing gel was pre-run for 1 h at a constant power of 65 W, in 1x TBE buffer (90 mM Tris-borate, 2 mM EDTA). A 4 µl volume of each sequencing reaction was loaded into the sample wells and electrophoresed at a constant power of 55 W for 2-6 h. The gel was then soaked in 5 % (v/v) acetic acid, 15 % (v/v) methanol for 20 min, dried at 80°C under vacuum, and exposed to X-ray film for 1-3 days. The X-ray film was then developed using an Amersham hyperprocessor.

CHAPTER 3

PURIFICATION OF CITRATE SYNTHASE FROM *PYROCOCCUS FURIOSUS*

3.0 INTRODUCTION

In order to achieve the project aim of cloning and sequencing the *P. furiosus* citrate synthase gene, it was essential to determine the amino acid sequence of a portion of the enzyme to which an oligonucleotide probe could be designed. This would then be used to analyse *P. furiosus* genomic DNA with the objective of cloning the citrate synthase gene. This chapter describes the preparation of cell extracts, the purification of the enzyme and the determination of the amino-terminal sequence.

3.1 MATERIALS

P. furiosus cell paste was generously donated by Dr. R. Sharp, Centre for Applied Microbiological Research, Porton Down, UK. The chromatographic matrix, Matrex™ gel Red A was obtained from Amicon Ltd., Stonehouse, UK. and a pre-packed Mono Q® ion exchange column from Pharmacia. Other materials, required for general protein methods, are detailed in Section 2.1.1.

3.2 METHODS

3.2.1 Preparation of *P. furiosus* cell extracts

P. furiosus cell paste (3-5 g wet weight) was resuspended in a 20 ml volume of buffer A (20 mM Tris-HCl pH 8.0, 25 mM NaCl, 2 mM EDTA) and lysed by sonication, on ice, with 4 x 30 s bursts at 140 W. Cell debris was then removed by centrifugation at 16,000 x *g* for 30 min.

3.2.2 Determination of the amino-terminal sequence of the *P. furiosus* citrate synthase

A Matrex gel Red A column, of volume 10 ml, was attached to a Minipuls 2, peristaltic pump (Anachem) to maintain a flow rate of 1 ml/min. After equilibrating the column in buffer A (Section 3.2.1), the *P. furiosus* cell extract was applied and the unbound material washed through the column with a 150 ml volume of this buffer. The *P. furiosus* citrate synthase was then eluted with a 10 ml volume of buffer A, supplemented with 5 mM oxaloacetate and 1 mM CoA (pH 8.0 (NaOH)).

The eluate was then diluted 4-fold with 25 mM diethanolamine-HCl pH 9.2 containing EDTA to a concentration of 2 mM, and applied to a Mono Q anion exchange column previously equilibrated in this buffer. A flow rate of 1 ml/min and the collection of fractions were controlled by a Fast Performance Liquid Chromatography (FPLC®) system (Pharmacia). Elution was achieved by a linear gradient of 0-0.35 M NaCl over a volume of 35 ml. Fractions were assayed for citrate synthase activity (Section 2.1.2) and the active fractions pooled. The protein in the pooled fractions was then precipitated with 10 % (w/v) trichloroacetic acid (TCA), analysed by SDS-PAGE, and electroblotted onto a PVDF membrane as described in Sections 2.1.4 and 2.1.5, respectively. Protein bands that required microsequencing were cut out, air dried and sequenced as described in Section 2.1.6.

3.3 RESULTS

3.3.1 Citrate synthase activity in *P. furiosus* cell extracts

Citrate synthase was measured at 55°C by the method of Srere *et al.* [1963] as described in Section 2.1.2. The specific activity of citrate synthase in *P. furiosus* cell extracts was determined as 5-8 nmoles/min/mg protein at 55°C.

3.3.2 Purification of citrate synthase from *P. furiosus*

A combination of dye-ligand and anion exchange chromatography was used to purify citrate synthase from cell extracts of *P. furiosus*. The enzyme bound to the dye-ligand matrix, Matrex gel Red A and was biospecifically eluted with buffer supplemented with oxaloacetate and CoA, resulting in an 84-fold increase in enzyme purity, as shown in Table 3.1. On applying this eluate to the Mono Q anion exchange column, the enzyme eluted in a symmetrical peak at 0.16 M NaCl, as shown in Figure 3.1. A 432-fold purification resulted in the observation of three major protein bands on a SDS-PAGE gel under reducing conditions. Thus, the enzyme must constitute less than 0.2 % of the total protein in *P. furiosus* cell extracts.

	Volume (ml)	Total Activity (nmoles/min)	Total Protein (mg)	Specific Activity (nmoles/min/mg)	Yield (%)	Purification
Cell extract	32.0	1700	216	8	100	-
Matrex gel Red A	4.4	1413	2.1	673	83	84
Mono Q	4.0	242	0.07	3460	14.0	432

Table 3.1 The purification of citrate synthase from *P. furiosus* cell extracts

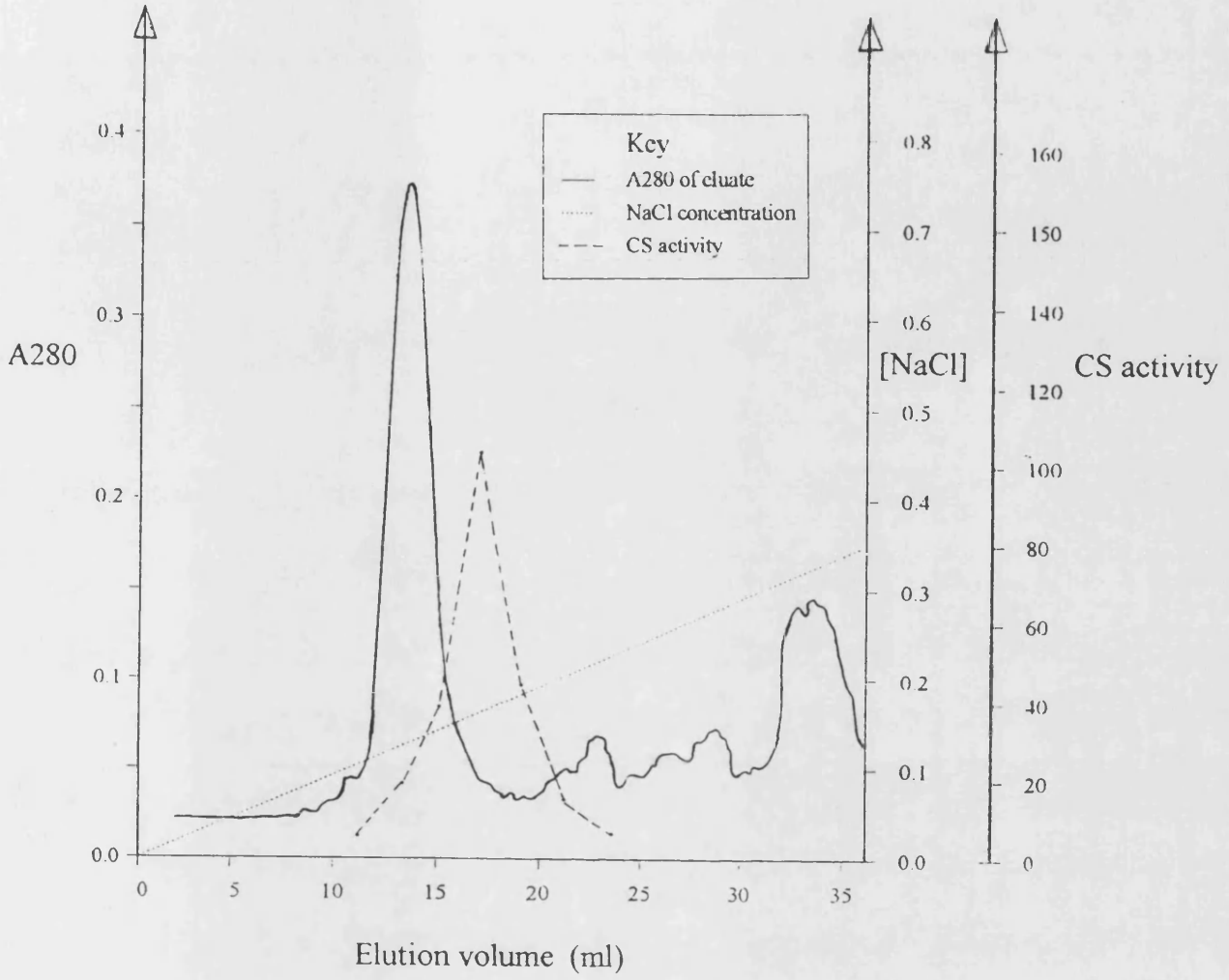


Figure 3.1 Anion exchange chromatography of the *P. furiosus* citrate synthase.

A preparation of *P. furiosus* citrate synthase, partially purified by affinity-dye chromatography, was applied to a Mono Q anion exchange column and eluted with a linear gradient of 0 - 0.35 M NaCl. *P. furiosus* citrate synthase eluted with 0.16 M NaCl. The units of citrate synthase activity are nmoles/min/mg protein. The molar amount of NaCl in the elution buffer is indicated by the dotted line.

3.3.3 Amino-terminal sequence of the three major bands in the *P. furiosus* citrate synthase preparation

Analysis of the citrate synthase preparation by SDS-PAGE demonstrated that it consisted of 3 major and 2 minor protein bands (Figure 3.2). The three major protein bands were subjected to microsequencing (Figure 3.3). Band **a** gave a major and a minor signal with each cycle of Edman degradation. The sequence of each band was compared to those in the protein databases using GCG Sequence Analysis Software (University of Wisconsin). These are discussed below (Section 3.4).

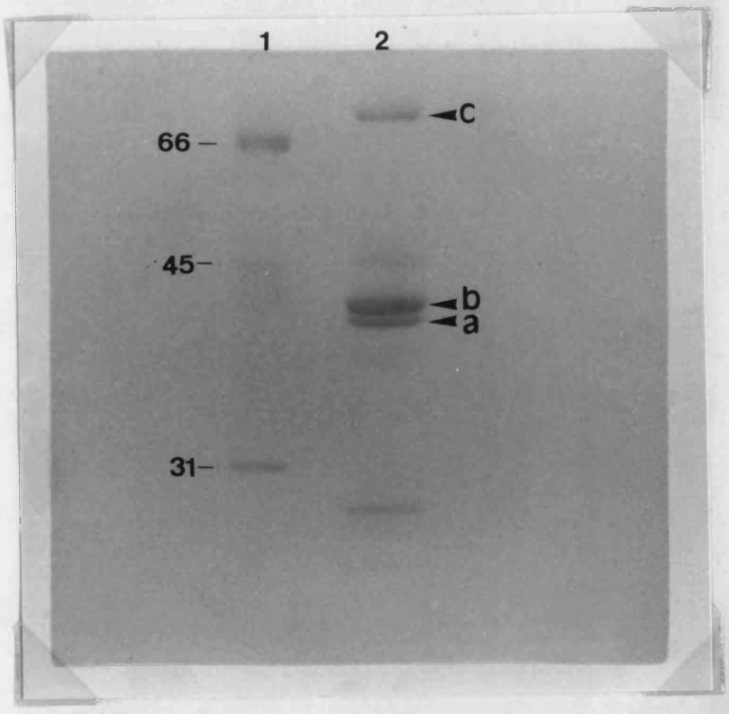


Figure 3.2 PVDF blot of a preparation of *P. furiosus* citrate synthase.

Lane 1: Low molecular weight standards
The M_r in kDa of each protein standard is indicated in the left hand margin.
Lane 2: *P. furiosus* citrate synthase preparation
The arrows indicate protein bands that were subjected to microsequencing.

Band **a**

Major sequence G L E D V Y I D Q T N I(Y)Y I D G K E G K L Y(Y)R(S)

Minor sequence M E K Y L A K V L I D(Q). I(D)Q(T)N . . Y

Band **b** S V I K M T D L D L A G K R V F I R A D . L N V P V K
D(G)K V

Band **c** D F D E Q E Q A R G I T I N A A N V S M V(H)(N) Y

Figure 3.3 Amino-terminal sequences of the three major protein bands in the *P. furiosus* citrate synthase preparation. The brackets denote a tentative residue assignment.

3.4 DISCUSSION

3.4.1 Citrate synthase activity in *P. furiosus* cell extracts

As *P. furiosus* is an anaerobic microorganism, it is likely to use the citric acid cycle solely for biosynthetic purposes. This was reflected in the extremely low specific activity of citrate synthase (5-8 nmole/min/mg) in *P. furiosus* cell extracts. For comparison, the citrate synthase in cell extracts of an aerobic, archaeal hyperthermophile, *S. solfataricus*, has a specific activity of 370 nmoles/min/mg at 55°C [M. J. Danson, personal communication]. It should be noted, that due to the thermolability of oxaloacetate, all assays were performed at 55°C which is sub-optimal for the *P. furiosus* enzyme.

These complications necessitated an affinity purification step, namely Matrex gel Red A. The matrix consists of agarose beads coated with a triazine dye, Procion Red, which mimics the adenine moiety found in various substrates and co-factors such as CoA and NADH. The purification protocol resulted in three major bands, which were subjected to amino-terminal sequencing.

3.4.2 Band a - a putative citrate synthase

Band **a** (41 kDa) gave a major and a minor signal implying that there may be two proteins present in the 41 kDa protein band. However, as shown by an alignment of sequence at the amino-terminus of known citrate synthases (Figure 3.4), the major sequence exhibited 64 % identity to the *S. solfataricus* enzyme [Lill *et al.*, 1992] and 55 % to that of *T. acidophilum* [Sutherland *et al.*, 1990]. This indicated that the major component of the 41 kDa band was indeed the *P. furiosus* citrate synthase and thus, provided information to assist the cloning of the citrate synthase gene.

On obtaining the entire gene sequence of the *P. furiosus* citrate synthase gene (Chapters 4 and 5), the position of the translation start codon was established. Suprisingly, the start codon was located 27 bp, in the 5' direction, from the codon for the Gly residue that was inferred by protein microsequencing to be the amino-terminal residue. Further examination of the 9 amino acids, now predicted to constitute the amino-terminus of enzyme, revealed that 6 residues corresponded exactly to the minor amino acid sequence determined by protein microsequencing, as shown in Figure 3.5. This indicates that the 41 kDa protein band comprises a single protein species that has been fragmented by cellular proteases during either the harvesting of the cells or the purification of the protein. Such degradation of the enzyme may also explain the low citrate synthase activity measured in the cell extract and during the purification. Alternatively, the two amino acid sequences may result from *in vivo* processing of the enzyme.

```
                Chicken  .......... .......ASS TNLKDVLAAL IPKEQARIKT FRQQHGGTAL GQITV.DMSY GGMRGMKGLV YETSVLDPDE G.IRFRGFSI
                    Pig  .......... .......ASS TNLKDILADL IPKEQARIKT FRQQHGNTVV GQITV.DMMY GGMRGMKGLV YETSVLDPDE G.IRFRGYSI
             C. elegans  .......... .....SAEGS TNLKEVLSKK IPAHNAKVKS FRTEHGSTVV QNVNI.DMIY GGMRSMKGMV TETSVLDPEE G.IRFRGYSI
            Yeast(glyo)  .......... .......SQE KTLKERFSEI YPIHAQDVRQ FVKEHGKTKI SDVLL.EQVY GGMRGIPGSV WEGSVLDPED G.IRFRGRTI
            Yeast(mito)  .......... .......ASE QTLKERFAEI IPAKAQEIKK FKKEHGKTVI GEVLLEEQAY GGMRGIKGLV WEGSVLDPEE G.IRFRGRTI
             N. crassa   .......... ......SSKT QTLKERFAEL LPENIEKIKA LRKEHGSKVV DKVTL.DQVY GGARGIKCLV WEGSVLDAEE G.IRFRGKTI
          Te. thermophila .......... ........SQ TNLKKVIAEI IPQKQAELKE VKEKYGDKVV GQYTV.KQVI GGMRGMKGLM SDLSRCDPYQ G.IIFRGYTI

             A. anitratum ....SEATGK KAVLHLDGKE .IELPIYSGT LGPDVIDVKD VLAS.GHFTF DP........ .GFMATASCE SKITFIDGDK GILLHRGYPI
           Ps. aeruginosa .......ADK KAQLIIEGSA PVELPVLSGT MGPDVVDVRG LTAT.GHFTF DP........ .GFMSTASCE SKITYIDGDK GVLLHRGYPI
                E. coli  .......ADT KAKLTLNGDT AVELDVLKGT LGQDVIDIRT LGSK.GVFTF DP........ .GFTSTASCE SKITFIDGDE GILLHRGFPI
             Ac. acetii  SASQKEGKLS TATISVDGKS A.EMPVLSGT LGPDVIDIRK LPAQLGVFTF DP........ .GYGETAACN SKITFIDGDK GVLLHRGYPI
             Co. burnetii .......SNR KAKLSFENQS .VEFPIYSPT LGKDVIDVKT L.GNHGAYAL DV........ .GFYSTAACE SKITFIDGEK GILLYRGYPI
            R. prowazekii .TNGNNNNLE FAELKIRGK. LFKLPILKAS IGKDVIDISR VSAEADYFTY DP........ .GFMSTASCQ STITYIDGDK GILWYRGYDI

          B. subtilis (CitZ) .......... .......... .......... .......... .........TA TR........ .GLEGVVATT SSVSSI..ID DTLTYVGYDI
          B. subtilis (CitA) .......... .......... .......... .......... .........V HY........ .GLKGITCVE TSISHIDGEK GRLIYRGHHA
B. coagulans (strain C4) .......... .......... .......... .......... ....VNTNQF IP........ .GLEGVIASE TKISFLDTVN SEIVIKGYDL
             M. smegmatis .......... .......... .......... .........T TATESEAPRI HK........ .GLAGVVVDT TAISKVVPET NSLTYRGYPV

           T. acidophilum .......... .......... .......... .......... ....PETEEI SK........ .GLEDVNIKW TRLTTIDGNK GILRYGGYSV
           S. solfataricus .......... .......... .......... .......... .......... .......... .GLEDVKIKS TSLTYIDGVN GVLRY
                P. furiosus .......... .......... .......... .......... .......... .......... .GLEDVYIDQ TNICYIDGKE GKLYYRGYSV
                H. volcanii .......... .......... .......... .......... ......XGEL KR........ .GLEGVLVAE SKLSFIDGDA GQLVYXGYDI
```

Figure 3.4 Alignment of sequence at the amino-terminus of known citrate synthases with the sequence determined by microsequencing the enzyme from *Sulfolobus solfataricus* [Lill *et al.*, 1992], *P. furiosus* and *Haloferax volcanii* [James, 1991].

The alignment was obtained using PILEUP (University of Wisconsin GCG, Version 7). The sequence from the citrate synthase of *S. solfataricus* was determined after limited proteolysis of the purified enzyme [Lill *et al.*, 1992]. Yeast (glyo) and (mito) refer to the glyoxysomal and the mitochondrial citrate synthases, respectively. The sequence shown from the *P. furiosus* enzyme is the major signal obtained on microsequencing the purified enzyme. Upon elucidation of the gene sequence, an additional eight residues were found to constitute the amino-terminus (Figure 3.5).

As shown in Figure 3.5, the amino acid sequence, derived from the nucleotide sequence, did not match the minor sequencing signal exactly, even though the Met codon illustrated here was the only candidate for a start codon that was in frame with the rest of the coding region. Although it is impossible to explain this discrepancy, the expression studies described in Chapter 6 demonstrate that the coding region elucidated from the cloning procedures encoded a functional, hyperthermostable protein.

Gene Sequence	ATG	AAT	ACG	GAA	AAA	TAC	CTT	GCT	AAA	GGT	CTA	GAA
Derived Sequence	M	N	T	E	K	Y	L	A	K	G	L	E
41 kDa Major Sequence										G	L	E
41 kDa Minor Sequence			M	E	K	Y	L	A	K	V	L	I

Gene Sequence	GAT	GTC	TAC	ATA	GAT	CAA	ACA	AAT	ATA	TGC	TAT	ATC
Derived Sequence	D	V	Y	I	D	Q	T	N	I	C	Y	I
41 kDa Major Sequence	D	V	Y	I	D	Q	T	N	I	X	Y	I
41 kDa Minor Sequence	D	(Q)	(Y)	I	(D)	Q	(T)	N	X	X	Y	

Gene Sequence	GAT	GGA	AAA	GAG	GGA	AAG	CTA	TAC	TAC	AGA	GGA
Derived Sequence	D	G	K	E	G	K	L	Y	Y	R	G
41 kDa Major Sequence	D	G	K	E	G	K	L	Y	(Y)	R	(S)

Figure 3.5 Alignment of the nucleotide-derived, amino-terminal sequence of citrate synthase from *P. furiosus* (Chapter 5) with the major and minor amino-terminal signals from the 41 kDa protein band in the purified citrate synthase preparation (band **a**). The brackets indicate a tentative residue assignment. X represents an unknown residue.

The amino-terminal sequence demonstrates that, in common with the enzyme from *T. acidophilum*, the *P. furiosus* enzyme is approximately 35 residues shorter than the eukaryal and Gram-negative bacterial citrate synthases sequenced to date. The absence of these residues is a feature exhibited by all known archaeal citrate synthases and by the enzyme from Gram-positive Bacteria. It is tempting to postulate that the smaller, more compact structure of these enzymes contributes to enzyme stability, but it is likely that possession of a recessed amino-terminus is a phylogenetically-derived characteristic. Certain Gram-negative Bacteria (*E. coli*, *Ps. aeruginosa*) have been shown to possess two citrate synthases; the predominant, hexameric form and a second, dimeric enzyme [Patton *et al.*, 1993; Anderson *et al.*, 1993]. The amino-terminal sequences of the *E. coli* and *Ps. aeruginosa* second citrate synthases have been determined. Both sequences align with the archaeal and Gram-positive bacterial citrate synthases in that they possess a recessed amino-terminus. In addition, the 'small', Gram-negative bacterial citrate synthases exhibit greater sequence identity to the archaeal enzymes than they do to their hexameric counterparts. This may imply that the 'small' citrate synthase may represent a gene that has duplicated and diverged to form the complex hexameric enzyme

3.4.3 Band b - *E. coli* 3-phosphoglycerate kinase

Surprisingly, band **b** (45 kDa) showed 100 % identity with 3-phosphoglycerate kinase (PGK) from *E. coli* [Alefounder and Perham, 1989] indicating that contamination of the citrate synthase preparation had occurred. In order to investigate the source of the contamination, the cell extract preparation and purification protocol were repeated in the absence of *P. furiosus* cells. As no activity could be detected at any step, the contamination was assumed to have occurred during cell harvesting. With further investigation, it was found that the cells were harvested from a continuous culture of *P. furiosus* and remained at room temperature until centrifugation at the end of the day. As mentioned earlier, the extremely low levels of citrate synthase within *P. furiosus* necessitated an affinity-dye

purification step, which also purifed the *E. coli* PGK. As purification by this method is very efficient, only a small amount of *E. coli* contamination was necessary to concentrate the contaminating protein. As the cell paste was contaminated, it was likely that *E. coli* citrate synthase had also bound to the Matrex gel Red A and co-eluted with the *P. furiosus* enzyme. The presence of two citrate synthases in the affinity purification eluate would explain the large decrease in yield during the Mono Q anion exchange (Table 3.1) which separated the two enzymes.

3.4.4 Band c - a putative elongation factor

The amino-terminal sequence of band **c** (75 kDa) exhibited 79 % identity to amino acids 60-80 of elongation factor 2 (EF-2) from *T. acidophilum* [Pechmann *et al*, 1991], as shown in Figure 3.6. EF-2 is analogous to bacterial elongation factor G (EF-G) and eukaryal eEF-2 which bind to the ribosome in the presence of GTP and effect the translocation of the polypeptide chain from the A to the P site of the ribosome in preparation for the next, incoming amino acyl-tRNA. EF-2 has been cloned and sequenced from a number of archaeal species including *Methanococcus vannielli* [Lechner *et al.*, 1988], *Halobacterium halobium* [Itoh, 1989], *T. acidophilum* [Pechmann *et al.*,1991] and *Sulfolobus acidocaldarius* [Schröder and Klink, 1991]. The nucleotide sequence of EF-2 of *P. furiosus* is not yet available. However, aligning the amino-terminal sequence of band **c** with the sequences of EF-2 from the archaeal species listed above (Figure 3.6), indicates that the *P. furiosus* elongation factor is either 57-59 residues shorter than its archaeal counterparts, or, like the *P. furiosus* citrate synthase, has been degraded during cell harvesting.

```
                1                                                   50
Me. vannielli   MGRRAKMVEK VKSLMETHDQ IRNMGICAHI AHGKTTLSDN LLAGAGMISK
H. halobium     MGRRKKIVEQ CERLMDNPEQ IRNIAIAAHV DHGKTTLTDN LLAGAGMISE
S. acido.       ..PRYKTVEQ VLSLMKDVTR VRNIGIIAHV DHGKTTTSDT LLAASGIISQ
T. acidophilum  MGRKEDNIEK ALKIVEHTEL IRNIGIVAHI DHGKTTLSDN LIAGAGMMSE
P. furiosus     .......... .......... .......... .......... ..........

                 51
Me. vannielli   DLAGDQLALD FDEEEAARGI TIYAANVSMV HEYNGKEYLI NL
H. halobium     DTAGQQLAMD TEEDEQERGI TIDAANVSMT HEYEGDDHLI NL
S. acido.       KVAGEALALD YLSVEQQRGI TVKAANISLY HEIDGKGYVI NL
T. acidophilum  ELAGKQLVLD YDEQEQARGI TINAAVASMV HTFQGKEYLI NL
P. furiosus     .........D FDEQEQARGI TINAANVSMV HNY....... ..
```

Figure 3.7 Alignment of the amino acid sequence derived from the 70 kDa protein which co-purified with *P. furiosus* citrate synthase with the amino-terminal sequences of EF-2 from *Methanococcus vannielli* [Lechner *et al.*, 1988], *Halobacterium halobium* [Itoh, 1989], *Thermoplasma acidophilum* [Pechmann *et al.*, 1991] and *Sulfolobus acidocaldarius* (*S. acido.*) [Schröder and Klink, 1991].

3.4.5 Conclusions

The amino-terminal degradation of the *P. furiosus* citrate synthase and the possible cleavage of EF-2 indicates proteolytic activity in the *P. furiosus* cell extracts despite being stored at 4°C. The regulation of proteolytic activity in this organism has been studied by Snowden and co-workers [Snowden *et al.*, 1992] who found that *P. furiosus* was unable to grow on a mixture of 20 amino acids, unless supplemented with tryptone or yeast extract, even if maltose was added as a carbon and energy source. However, it was able to utilise proteins and peptides as the sole carbon, nitrogen and energy source. In this situation, the level of proteolytic activity, as measured by azocasein hydrolysis, was higher than that found in conventional media. Therefore, for future reference, it is vital to include protease inhibitors when purifying enzymes from this organism.

Overall, despite complications involving low enzyme levels, enzyme degradation and contaminated starting material, citrate synthase was purified from cell extracts of *P. furiosus* and yielded a small region of sequence that showed homology to the citrate synthase of *S. solfataricus* and *T. acidophilum*. The following chapter describes the design of an oligonucleotide to a region of this sequence and its use as a primer to amplify a fragment of the *P. furiosus* citrate synthase gene.

CHAPTER 4

CLONING AND SEQUENCING THE AMINO-TERMINAL PORTION OF THE *PYROCOCCUS FURIOSUS* CITRATE SYNTHASE GENE

4.0 INTRODUCTION

Cloning the citrate synthase gene from *P. furiosus* is essential to this investigation into the molecular basis of protein thermostability. It will also yield new sequence information to contribute to a protein-derived, phylogenetic analysis (Section 1.2), and the expression of the cloned gene in a mesophilic host such as *E. coli* JM105 will supply adequate quantities of enzyme for thermostability and X-ray crystallographic studies. The expression of the citrate synthase gene is imperative for the advanced study of this enzyme as the level of citrate synthase in *P. furiosus* cell extracts is extremely low (Section 3.4.1).

The previous chapter has described the purification of citrate synthase from cell extracts of *P. furiosus* and the subsequent determination of amino acid sequence in the region of the amino-terminus. The sequence obtained exhibited 55 % identity to the *T. acidophilum* citrate synthase, providing key information for the design of an oligonucleotide to instigate the cloning of the citrate synthase gene. This chapter details the design of such an oligonucleotide and its use as a primer to amplify a portion of the citrate synthase coding region by the polymerase chain reaction (PCR).

PCR is a powerful technique whereby a desired nucleotide sequence can be specifically amplified from a complex target sequence such as total, genomic DNA [Saiki *et al.*, 1985; Mullis *et al.*, 1986; Mullis and Faloona, 1987]. The technique relies on the specific annealing of two oligonucleotide primers to opposite strands of the target DNA, such that their 3' termini point towards each other across the desired sequence. Repeated

cycles of target denaturation, primer annealing and enzyme-catalysed primer extension, using a thermostable DNA polymerase, then effect the amplification of the desired region.

A multiple sequence alignment of all known citrate synthases was used to pinpoint a region of high sequence identity to which an antisense, redundant oligonucleotide was designed. This was then used in conjunction with the oligonucleotide designed to amino acid sequence at the amino-terminus of the enzyme to amplify the intervening portion of the citrate synthase gene. This chapter details the design of the primers and the subsequent amplification, cloning and sequencing of a fragment of the citrate synthase gene.

4.1 MATERIALS

P. furiosus genomic DNA was generously donated by Dr R. Sharp, Centre for Applied Microbiological Research, Porton Down, UK. Suppliers of general reagents for molecular biology methods are detailed in Section 2.2.1. *E. coli* strains and plasmids are detailed in Section 2.2.2, as are conditions for the culture of *E. coli*.

4.2 METHODS

4.2.1 Design and synthesis of the forward amplification primer

An oligonucleotide probe was designed to the first 12 residues of the major amino-terminal sequence of the 41 kDa protein, purified in the previous chapter (Figure 3.3). To assist oligonucleotide design, codon bias in *P. furiosus* was assessed by the construction of a codon usage table (Table 4.1) from the gene sequences of two *P. woesei* proteins, GAPDH [Zwickl *et al.*, 1990] and elongation factor 1α [Creti *et al.*, 1991]. No *P. furiosus* sequences were available at the time of oligonucleotide design. A 35 base (35-mer), single-stranded oligonucleotide, denoted PfCS(N), was designed (Figure 4.1). The oligonucleotide was synthesised and purified as described in Section 2.2.9.

Amino acid	Codon	Fraction	Amino acid	Codon	Fraction
Ala	GCG	0.02	Leu	TTG	0.09
Ala	GCA	0.32	Leu	TTA	0.07
Ala	GCT	0.44	Leu	CTG	0.00
Ala	GCC	0.23	Leu	CTA	0.12
			Leu	CTT	0.35
Arg	AGG	0.41	Leu	CTC	0.37
Arg	AGA	0.59			
Arg	CGG	0.00	Lys	AAG	0.76
Arg	CGA	0.00	Lys	AAA	0.24
Arg	CGT	0.00			
Arg	CGC	0.00	Met	ATG	1.00
Asn	AAT	0.26	Pro	CCG	0.05
Asn	AAC	0.74	Pro	CCA	0.73
			Pro	CCT	0.09
Asp	GAT	0.51	Pro	CCC	0.14
Asp	GAC	0.49			
			Phe	TTT	0.31
Cys	TGT	1.00	Phe	TTC	0.69
Cys	TGC	0.00			
			Ser	AGT	0.13
Gln	CAG	0.57	Ser	AGC	0.43
Gln	CAA	0.43	Ser	TCG	0.00
			Ser	TCA	0.17
Glu	GAG	0.49	Ser	TCT	0.13
Glu	GAA	0.51	Ser	TCC	0.13
Gly	GGG	0.06	Thr	ACG	0.02
Gly	GGA	0.51	Thr	ACA	0.49
Gly	GGT	0.36	Thr	ACT	0.29
Gly	GGC	0.08	Thr	ACC	0.20
His	CAT	0.19	Trp	TGG	1.00
His	CAC	0.81			
			Tyr	TAT	0.32
Ile	ATA	0.26	Tyr	TAC	0.68
Ile	ATT	0.40			
Ile	ATC	0.33	Val	GTG	0.08
			Val	GTA	0.38
End	TGA	0.50	Val	GTT	0.37
	TAA	0.50	Val	GTC	0.17
	TAG	0.00			

Table 4.1 Codon usage table derived from the nucleotide sequences of glyceraldehyde-3-phosphate dehydrogenase and elongation factor 1α from *P. woesei* [Zwickl *et al.*, 1990; Creti *et al.*, 1991]

Amino acid	G	L	E	D	V	Y	I	D	G	T	N	I
Nucleotide	GGA	CTT	GAA	GAT	GTA	TAC	ATA	GAC	CAG	ACA	AAC	AT
5'-3'	T	C	G	C	T	T	T	T	A	T	T	
							C					

Figure 4.1 Nucleotide sequence of the forward amplification primer, PfCS(N)

4.2.2 Design and synthesis of the reverse amplification primer

A multiple sequence alignment of all known citrate synthases (Figure 7.2) revealed a region of eight residues, corresponding to residues 258-265 of the *T. acidophilum* enzyme, which are highly conserved (Figure 4.2, residues 353-360 of Figure 7.2)). Of these eight residues, three are conserved throughout all species whereas the remaining five are conserved between the *T. acidophilum* citrate synthase and the enzyme from bacterial sources. Therefore, as it is probable that this sequence, or one very similar, will be found in the *P. furiosus* enzyme, this region constitutes an ideal template for the design of a redundant oligonucleotide. In addition, the amino acid residues of this region are encoded by a limited number of nucleotide codons, thereby lowering the redundancy of the oligonucleotide.

A 23 base (23-mer), redundant oligonucleotide, denoted PfCS(P), was designed to the sequence M G F G H R V Y using the codon usage table detailed above (Table 4.1, Section 4.2.1), to assess codon bias. However, in order for PfCS(P) to function with PfCS(N) as a PCR primer, the antisense strand of the predicted nucleotide sequence was required. The oligonucleotide was synthesised and purified as described in Section 2.2.9.

Chicken	NSGRVVPGYGHAVLRKT.D
Pig	NSGRVVPGYGHAVLRKT.D
Yeast	NSGRVIPGYGHAVLRKT.D
B. coagulans (strain C4)	KE..KI**MGFGHRVY**MKKMD
Ps. aeruginosa	NDPFKL**MGFGHRVY**.KNFD
A. anitratum	E..VKL**MGFGHRVY**.KNFD
E. coli	NDSFRL**MGFGHRVY**.KNYD
T. acidophilum	NGKKRL**MGFGHRVY**.KTYD

Figure 4.2 Multiple sequence alignment of a region of citrate synthase that is highly conserved between all known citrate synthases.

Eight residues of this region (indicated by bold font) were used as a template for the design of a redundant, antisense oligonucleotide, PfCS(P). These residues correspond to amino acids 258-265 of the *T. acidophilum* enzyme.

Amino acid	M	G	F	G	H	R	V	Y
Nucleotide sequence (antisense)	TAC	CCT	AAA	CCT	GTA	TCC	CAT	AT
3'-5'		A	G	A	G	T	A	

Figure 4.3 Nucleotide sequence of the reverse amplification primer, PfCS(P)

4.2.3 Amplification of a fragment of the *P. furiosus* citrate synthase gene

A 100 ng quantity of *P. furiosus* genomic DNA was added to 50 pmoles of each primer and 75 nmoles of each deoxynucleoside 5' triphosphate (dATP, dCTP, dGTP, dTTP) in a 0.1 ml volume of 1x PCR reaction buffer (supplied by the manufacturer) containing 0.1 mg BSA/ml. The reaction was heated to 94°C for 10 min to denature the DNA before the addition of 2.5 U *Taq* DNA polymerase. To amplify the gene, the following cycle was repeated 25 times: 94°C for 1.25 min, 55°C for 0.5 min, 72°C for 2 min. After the final cycle, the reaction was incubated at 72°C for 10 min to ensure that primer extension was complete. Reaction temperature and cycle programming were controlled by a Cetus DNA Thermal Cycler (Perkin-Elmer Ltd., Beaconsfield, UK.). The amplification programme was repeated with control samples in which either the target DNA, PfCS(N) or PfCS(P) was replaced with water.

4.2.4 Preparation of the amplification product for blunt-ended ligation into pUC18

The amplification product was treated with DNA polymerase I to 'end-fill' any single-stranded, nucleotide overhangs generated by the *Taq* DNA polymerase. The amplification product was separated from the PCR reaction components by agarose gel electrophoresis (Section 2.2.6) and purified from the agarose gel by the freeze-squeeze method (Section 2.2.7). Approximately 100 ng of the purified, amplification product was incubated for 5 min at 37°C in a 20 µl volume of 2x One-Phor-All buffer (Pharmacia) containing 9 U *E. coli* DNA polymerase I. The reaction mixture was then supplemented with 4 nmoles of each deoxynucleoside 5' triphosphate (dATP, dCTP, dGTP, dTTP) and the reaction incubated at 37°C for a further 30 min. The amplification product was then purified by increasing the reaction volume to 200 µl with distilled water, extracting twice with an equal volume of 25:24:1 (v:v:v) phenol:chloroform:isoamylalcohol and precipitating the DNA with ethanol (Section 2.2.4).

4.2.5 Ligation of the amplification product into pUC18-*Sma*I/BAP

The amplification product was treated with DNA polymerase I, as described above (Section 4.2.4), to ensure it possessed blunt-ended termini. It was then ligated into a plasmid vector, pUC18, that had previously been linearised with *Sma*I and treated with bacterial alkaline phosphatase (BAP) to minimise vector religation (pUC18-*Sma*I/BAP, Pharmacia). A 5 µl volume of the DNA polymerase I-treated, amplification product was incubated overnight at 15°C in 1x ligase buffer (supplied by manufacturer) containing 8 U T4 DNA ligase and 10 ng pUC18-*Sma*I/BAP in a total volume of 20 µl. To assess the degree of vector religation, a control ligation was prepared in which the amplification product was replaced by distilled water.

4.2.6 Transformation of *E. coli* XL1 Blue with the gene fragment ligated into pUC18

Competent *E. coli* XL1 Blue were prepared by the method of Hanahan [1985] and transformed with a 2 µl volume of the ligation reaction by the heat-shock method, as detailed in Section 2.2.14. Transformed cells were spread on LB-agar plates containing 100 µg ampicillin/ml (to select for the plasmid) and 0.8 mg IPTG (from a 200 mg/ml stock) and 0.8 mg X-gal (from a 20 mg/ml stock in dimethylformamide) to select for recombinants. The plates were then inverted and incubated overnight at 37°C.

To screen recombinants for the correct insert, white colonies were removed from the plate with a sterile toothpick and cultured overnight at 37°C in a 3 ml volume of 2x YT medium, supplemented with ampicillin to a concentration of 100 µg/ml. A 1.5 ml volume of each overnight culture was then used in a small-scale, plasmid preparation, detailed in Section 2.2.15. A 5 µl volume of the plasmid preparation was digested with 5 U *Bam*HI and 5 U *Kpn*I (Section 2.2.3) and analysed by agarose gel electrophoresis as described in Section 2.2.6. Recombinant PfCS(COL3) was selected for sequencing.

4.2.7 Nucleotide sequencing of the amplification product

The "maxiprep" method, detailed in Section 2.2.16, was used to obtain a large scale preparation of purified plasmid. The insert was then sequenced by the dideoxy-sequencing method of Sanger *et al.* [1977] as detailed in Section 2.2.17. The 760 bp insert was sequenced using the universal forward and reverse primers, which anneal to regions of the plasmid, and additional 17 base oligonucleotides designed to previously determined sequence (Figure 4.4). The sequencing primers were synthesised and purified as described in Section 2.2.9. The nucleotide sequence was then translated into amino acid sequence using STADEN PLUS Version 6, (Amersham International plc.) and aligned to the primary sequence of citrate synthase from *T. acidophilum* using the BESTFIT program of the GCG Sequence Analysis Software (University of Wisconsin GCG).

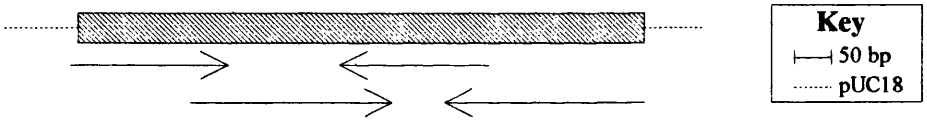


Figure 4.4 Strategy for the determination of the nucleotide sequence of a fragment of the *P. furiosus* citrate synthase gene, amplified from genomic DNA. The insert (760 bp) is represented by the hatched bar; the arrows represent the region of nucleotide sequence obtained from each primer.

4.3 RESULTS

4.3.1 Amplification of a fragment of the *P. furiosus* citrate synthase gene by PCR

A fragment of the gene encoding citrate synthase was amplified from *P. furiosus* genomic DNA using two redundant, oligonucleotide primers, PfCS(N) and PfCS(P) (Sections 4.2.1 and 4.2.2). The amplification procedure, described in Section 4.2.3, resulted in a single DNA product when analysed by agarose gel electrophoresis (Figure 4.5). By comparison with DNA size markers, consisting of *Hin*dIII-digested λ DNA, the amplification product was determined to be approximately 760 bp in length. This is comparable to the expected size (774 bp), estimated from the sequence of the *T. acidophilum* enzyme. Control reactions, in which either the target DNA, PfCS(N) or PfCS(P) was replaced by distilled water, did not yield any product.

4.3.2 Sequencing of the 760 bp fragment of the *P. furiosus* citrate synthase gene

The amplification product was successfully ligated into pUC18-*Sma*I/BAP and transformed into competent *E. coli* XL1 Blue as detailed in Sections 4.2.4 - 4.2.6. Transformants were screened for the recombinant plasmid by isolating plasmid DNA from each, digesting it with the restriction endonucleases *Bam*HI and *Kpn*I, and analysing the fragments by agarose gel electrophoresis. *Bam*HI and *Kpn*I possess single sites on pUC18 which flank the *Sma*I site and thereby, flank an insert ligated into the *Sma*I site. Figure 4.6 illustrates the nucleotide sequence of the 760 bp amplification product; an alignment of this sequence to the primary sequence of the *T. acidophilum* citrate synthase is shown in Figure 4.7.

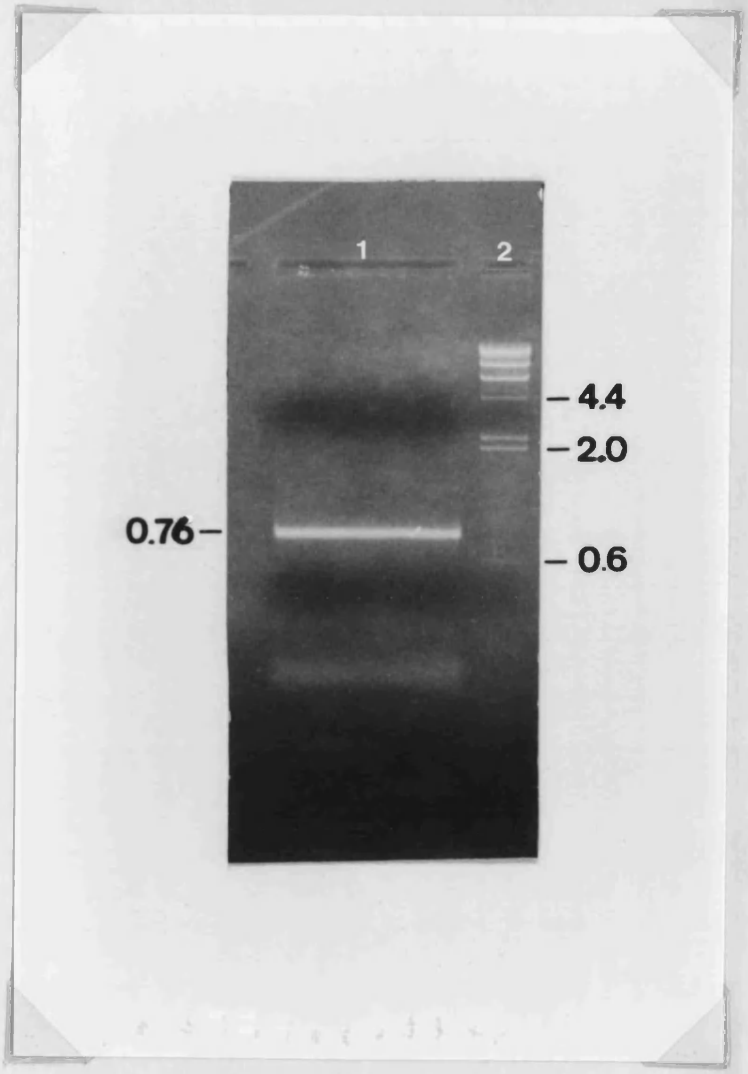


Figure 4.5 Preparative agarose gel of a fragment of the gene encoding citrate synthase that was amplified from *P. furiosus* genomic DNA

Lane 1: 760 bp fragment of the citrate synthase gene amplified from *P. furiosus* genomic DNA

Lane 2: DNA size markers - the length in kb of selected markers is indicated in the right hand margin

```
 L  E  D  V  Y  I  D  Q  T  N  I  C  Y  I  D  G  K  E  G  K   20
CTTGAGGACGTTTATATAGATCAGACTAATATATGCTATATCGATGGAAAAGAGGGAAAG  60

 L  Y  Y  R  G  Y  S  V  E  E  L  A  E  L  S  T  F  E  E  V   40
CTATACTACAGAGGATACAGTGTAGAAGAGCTAGCAGAGTTAAGCACGTTTGAGGAAGTG 120

 V  Y  L  L  W  W  G  K  L  P  S  L  S  E  L  E  N  F  K  K   60
GTGTATCTCCTCTGGTGGGGAAAATTACCATCCCTAAGTGAACTAGAGAATTTCAAAAAA 180

 E  L  A  K  S  R  G  L  P  K  E  V  I  E  I  M  E  A  L  P   80
GAGCTAGCAAAAAGTAGAGGCCTTCCAAAAGAAGTTATAGAAATAATGGAAGCACTACCT 240

 K  N  T  H  P  M  G  A  L  R  T  I  I  S  Y  L  G  N  I  D  100
AAAAATACTCACCCAATGGGCGCTTTAAGAACCATAATCTCATATCTAGGAAACATTGAC 300

 D  S  G  D  I  P  V  T  P  E  E  V  Y  R  I  G  I  S  V  T  120
GATAGTGGAGACATTCCAGTAACCCCTGAAGAAGTTTACAGGATAGGAATTAGTGTAACT 360

 A  K  I  P  T  I  V  A  N  W  Y  R  I  K  N  G  L  E  Y  V  140
GCAAAGATACCAACAATAGTAGCTAATTGGTATAGAATTAAGAATGGTCTCGAATATGTT 420

 P  P  K  E  K  L  S  H  A  A  N  F  L  Y  M  L  H  G  E  E  160
CCTCCAAAAGAAAAACTTAGTCATGCAGCAAATTTCCTATACATGCTCCACGGTGAAGAA 480

 P  P  K  E  W  E  K  A  M  D  V  A  L  I  L  Y  A  E  H  E  180
CCCCCCAAAGAATGGGAGAAAGCCATGGATGTGGCCCTAATTCTATATGCTGAACATGAA 540

 I  N  A  S  T  L  A  V  M  T  V  G  S  T  L  S  D  Y  Y  S  200
ATTAACGCATCGACTTTAGCAGTAATGACCGTTGGCTCTACTCTTAGCGACTACTATTCA 600

 A  I  L  A  G  I  G  A  L  K  G  P  I  H  G  G  A  V  E  E  220
GCCATCCTCGCGGGAATAGGAGCTTTAAAGGGCCCAATCCATGGAGGTGCAGTAGAGGAG 660

 A  I  K  Q  F  M  E  I  G  S  P  E  K  V  E  E  W  F  F  K  240
GCCATAAAACAGTTTATGGAAATAGGATCTCCCGAGAAAGTAGAAGAATGGTTCTTTAAG 720

 A  L  Q  Q  K  R  K  I  M  G  F  G  H   253
GCTCTTCAGCAGAAGAGAAAAATAATGGGTTTCGGACACA 760
```

Figure 4.6 Nucleotide sequence and derived amino acid sequence of the 760 bp fragment of the *P. furiosus* citrate synthase gene

```
P. furiosus         1 LEDVYIDQTNICYIDGKEGKLYYRGYSVEE.LAELSTFEEVVYLLWWGKL 49
                      |||| |  |    |||  | | | |||||   |     ||  ||   | |
T. acidophilum     10 LEDVNIKWTRLTTIDGNKGILRYGGYSVEDIIASGAQDEEIQYLFLYGNL 59

P. furiosus        50 PSLSELENFKKELAKSRGLPKEVIEIMEALPKNTHPMGALRTIISYLGNI 99
                      |   ||   |    |    |  ||     ||
T. acidophilum     60 PTEQELRKYKETVQKGYKIPDFVINAIRQLPRESDAVAMQMAAVAAMAAS 109

P. furiosus       100 DDSGDIPVTPEEVYRIGISVTAKIPTIVANWYRIKNGLEYVPPKEKLSHA 149
                                                |  | ||         ||   | |
T. acidophilum    110 E..TKFKWNKDTDRDVAAEMIGRMSAITVNVYRHIMNMPAELPKPSDSYA 157

P. furiosus       150 ANFLYMLHGEEPPKEWEKAMDVALILYAEHEINASTLAVMTVGSTLSDYY 199
                        ||    |    ||   ||  |||||  ||  ||| |     |||||  |
T. acidophilum    158 ESFLNAAFGRKATKEEIDAMNTALILYTDHEVPASTTAGLVAVSTLSDMY 207

P. furiosus       200 SAILAGIGALKGPIHGGAVEEAIKQFMEIGSPEKVEEWFFKA.LQQKRKI 248
                      | | |   ||||| |||| | || || ||  |  || ||       |
T. acidophilum    208 SGITAALAALKGPLHGGAAEAAIAQFDEIKDPAMVEKWFNDNIINGKKRL 257

P. furiosus       249 MGFGH 253
                      |||||
T. acidophilum    258 MGFGH 262
```

Figure 4.7 Alignment of the amino acid sequence encoded by the 760 bp fragment amplified from *P. furiosus* genomic DNA with the primary sequence of the *T. acidophilum* citrate synthase [Sutherland *et al.*, 1990]. The alignment was obtained using the BESTFIT program of the GCG Sequence Analysis Software (University of Wisconsin). Identical residues are indicated by |.

4.4 DISCUSSION

4.4.1 Amplification and cloning strategy

A fragment of the gene encoding citrate synthase was successfully amplified by PCR from *P. furiosus* genomic DNA using two redundant, oligonucleotide primers. The forward primer, PfCS(N), was designed to amino acid sequence determined from the purified gene product, whereas the reverse primer, PfCS(P), exploited sequence homology between known citrate synthases. Thus, the reverse primer was designed to inferred amino acid sequence. It is possible to obtain internal, amino acid information by the microsequencing of peptide fragments, generated by limited proteolysis or chemical cleavage of the purified protein. However, in this case, the citrate synthase of *P. furiosus* was not purified to homogeneity, or in sufficient amounts, to apply this approach.

A codon usage table, assembled from two proteins from a related species, *P. woesei*, indicated codon bias towards A-T rich, nucleotide sequences (Table 4.1). Consequently, PfCS(N) and PfCS(P) contain base multiplicity at the third position of each codon. If a primer that is not a perfect match is extended by *Taq* DNA polymerase, any mismatches occurring during the first cycle of amplification will become incorporated into the next. Thus, the nucleotide sequence of these regions of the amplification product must be regarded as influenced by the primers.

Prior to ligating the amplification product into pUC18-SmaI/BAP, it was treated with *E. coli* DNA polymerase I to remove any unpaired nucleotide 'overhangs' at the termini of the amplification product. This procedure has the disadvantage of removing a small number of nucleotides at each terminus. An alternative strategy is to engineer hexameric recognition sites for restriction endonucleases into the PCR primers which are incorporated into the amplification product during the first two cycles of the reaction.

Although this approach facilitates the subsequent ligation of the product into a standard cloning vector, it was not selected for this system. The oligonucleotide primers, used to amplify this fragment of the *P. furiosus* citrate synthase gene, were both designed to unclarified nucleotide sequence and in the case of PfCS(P), to unclarified, amino acid sequence. Thus, it was considered unwise to enhance the possibility of non-specific binding by using primer mismatches to engineer restriction sites into the product.

4.4.2 Nucleotide sequence of the 760 bp fragment amplified from *P. furiosus* genomic DNA

Translation of the 760 bp nucleotide sequence established that the PCR reaction had indeed amplified a region of open reading frame (Figure 4.6). On aligning the translated amino acid sequence of this reading frame to the primary sequence of the *T. acidophilum* citrate synthase, the sequences exhibited 38.2 % identity and 56.2 % similarity. In addition, an alignment of known citrate synthase sequences indicated that the *P. furiosus* sequence contained (Figure 7.2) one of the active site His residues (position 214 of Figure 4.6).

Overall, this application of PCR technology has provided some nucleotide sequence of the *P. furiosus* citrate synthase gene. This information was essential in obtaining the remainder of the gene sequence, as described in the following chapter. A detailed analysis of the complete sequence of the *P. furiosus* citrate synthase gene comprises Chapter 7.

CHAPTER 5

CLONING AND SEQUENCING THE REMAINDER OF THE *PYROCOCCUS FURIOSUS* CITRATE SYNTHASE GENE

5.0 INTRODUCTION

The previous chapter describes the amplification, cloning and sequencing of a 760 bp fragment of the *P. furiosus* citrate synthase gene, extending from the postulated amino-terminus to an internal region of high sequence conservation. The two oligonucleotide primers used to amplify this fragment, PfCS(N) and PfCS(P), bound specifically to *P. furiosus* genomic DNA at 55°C to produce a single product. This chapter details the use of these oligonucleotides as labelled probes, to locate the position of the gene in Southern blots [Southern, 1975] of restricted, *P. furiosus* genomic DNA.

Southern hybridisation identifies genomic fragments that contain all or part of the gene of interest, providing the oligonucleotide probe is specific for that gene. The standard approach to cloning a hybridising fragment [Sambrook *et al.*, 1989] is to repeat the appropriate genomic digestion and to purify, from an agarose gel, the DNA fragments corresponding to the estimated size of the hybridising fragment. The purified DNA is ligated into a linearised plasmid vector that possesses termini which are complementary to those of the genomic fragments, and the ligation reaction is used to transform competent *E. coli*. The resulting, size-restricted genomic library is then screened for the gene of interest.

Several attempts were made to prepare size-restricted genomic libraries of *P. furiosus* genomic fragments, without success. However, using information determined in the Southern hybridisation experiments, the hereto, undetermined regions of the citrate synthase gene were amplified from a circularised genomic fragment by inverse PCR. Inverse PCR is a variation of the standard amplification technique in which the target DNA consists of circularised genomic fragments [Ochman *et al.*, 1988; Silver and Keerikatte, 1989; Triglia *et al.*, 1988]. The amplification primers, designed to a region of known sequence within the target DNA, are orientated such that their 3' termini point away from each other, as illustrated in Figure 5.1. Consequently, primer extension results in the amplification of the larger segment of the circularised target, including the sequences that flank the region of known sequence.

In the case of the *P. furiosus* citrate synthase gene, the region of known sequence was the 760 bp gene fragment which was amplified, cloned and sequenced in the preceding chapter. The areas that flank this sequence comprise the portion of the gene encoding the carboxyl-terminus of the enzyme and the upstream, non-coding region. Thus, the cloning and sequencing of a gene fragment, amplified by inverse PCR, completed the determination of the citrate synthase gene sequence.

5.1 MATERIALS

P. furiosus genomic DNA was generously donated by Dr R. Sharp, Centre for Applied Microbiological Research, Porton Down, UK. Suppliers of reagents for molecular biology methods are detailed in Section 2.2.1. *E. coli* strains and plasmids are detailed in Section 2.2.2, as are conditions for the culture of *E. coli*.

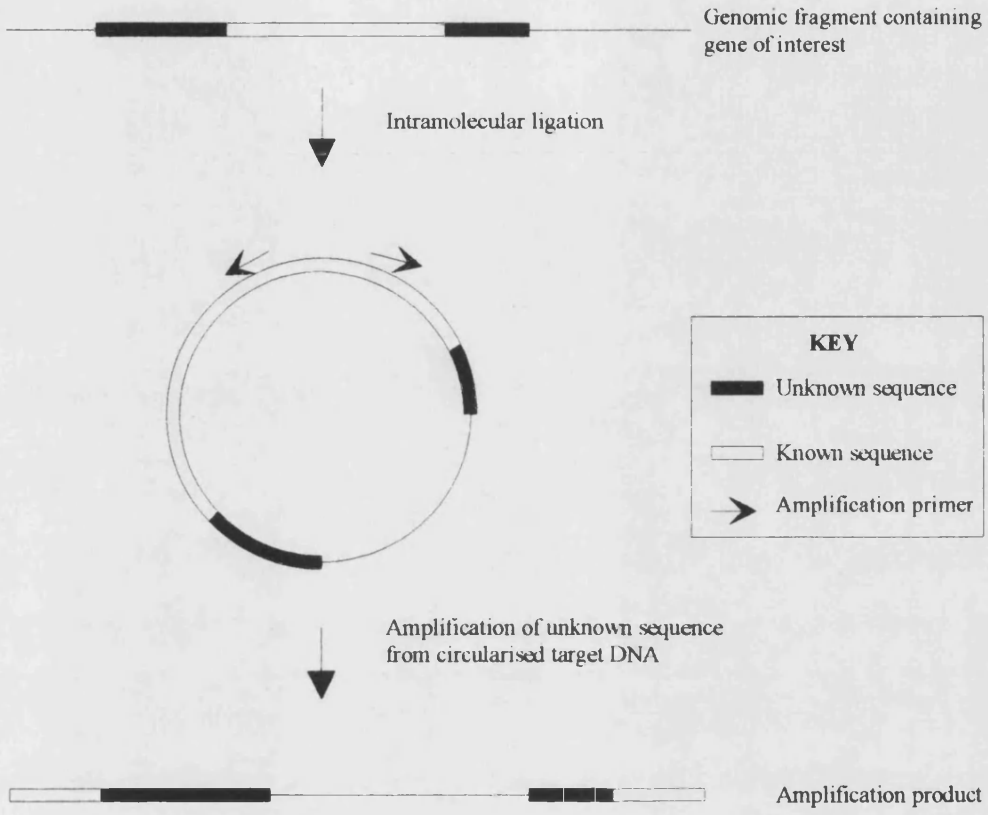


Figure 5.1 Strategy for an inverse PCR amplification reaction

Genomic DNA, fragmented using one restriction endonuclease, is ligated onto itself to form circular genomic fragments. Two amplification primers are designed to opposite strands of known sequence such that they are orientated away from each other. These are then used to amplify the regions flanking the known sequence.

5.2 METHODS

5.2.1 Analysis of *P. furiosus* genomic DNA by Southern hybridisation

10 μg quantities of *P. furiosus* genomic DNA were restricted with a number of restriction enzymes in both single and double digests (Section 2.2.5). The resulting DNA fragments were separated by agarose gel electrophoresis (Section 2.2.6) and the gel photographed under transillumination. The DNA fragments were then transferred, by the method of Southern [1975], onto a nylon membrane (Section 2.2.8).

Oligonucleotide PfCS(N) was end-labelled with [γ^{32}P] ATP (Section 2.2.10) and used to probe three membranes containing genomic DNA, digested with: (i) *Eco*RI, *Bam*HI, *Hin*dIII, *Pst*I, *Sac*I and *Xba*I, (ii) *Bam*HI+*Hin*dIII, *Pst*I+*Hin*dIII, *Xba*I+*Hin*dIII, *Bam*HI+*Pst*I, *Bam*HI+*Sac*I and *Bam*HI+*Xba*I, (iii) *Eco*RI, *Nde*I and *Eco*RI+*Nde*I, as detailed in Section 2.2.12. Oligonucleotide PfCS(P) was tailed with digoxygenin-11-dUTP/ATP (Section 2.2.11) and used to probe a Southern blot of genomic DNA digested with *Eco*RI, *Sph*I, *Eco*RI+*Sph*I, *Eco*RI+*Sal*I, *Eco*RI+*Bam*HI, *Eco*RI+*Pst*I, *Eco*RI+*Hin*dIII and *Eco*RI+*Xba*I, as detailed in Section 2.2.13. The sizes of the hybridising fragments were estimated from the polaroid photograph of the agarose gel taken before the Southern blotting procedure.

5.2.2 Attempted formation of a size-restricted gene library of *P. furiosus* genomic DNA

A 20 μg quantity of *P. furiosus* genomic DNA was restricted with *Eco*RI and *Hin*dIII as described in Section 2.2.5. Following agarose gel electrophoresis (Section 2.2.6), fragments of 1.58-1.62 kb were cut out of the gel and purified by the freeze-squeeze method (Section 2.2.7). In a similar manner, a 2 μg quantity of pUC18 was digested with *Hin*dIII and *Eco*RI and the linearised vector purified by the freeze-squeeze method,

following agarose gel electrophoresis to remove the intervening plasmid fragment. Equal amounts (approximately 50-250 ng) of vector and insert were incubated overnight, at 15°C, with 8 U T4 DNA ligase in a 20 µl volume of 1x ligase buffer (supplied by the manufacturer). Competent *E. coli* XL1 Blue were then transformed with a 2 µl volume of the ligation mix, as described in Section 2.2.14.

5.2.3 Amplification of the remainder of the citrate synthase gene by inverse PCR

A 32 µg quantity of *P. furiosus* genomic DNA was restricted with *Nde*I (Section 2.2.5) and the resulting DNA fragments separated by agarose gel electrophoresis (Section 2.2.6). Fragments of 1.58-1.62 kb were cut out of the gel and purified using the freeze-squeeze technique (Section 2.2.7), resulting in a final volume of 30 µl. A 4 µl volume of the purified genomic fragments was incubated at room temperature, for 4 h, with 8 U T4 DNA ligase in a 30 µl volume of 1x ligase buffer (supplied by the manufacturer). The low concentration of DNA in the ligation reaction favoured intramolecular ligation rather than intermolecular ligation and thus promoted the formation of circles of genomic DNA.

The oligonucleotide primers, selected for this amplification, were 17 nucleotides in length and designed to sequence determined in the previous chapter. The sequence of oligonucleotide PfCS(SEQ2) was identical to nucleotides 583-597 (inclusive) of the antisense strand of the citrate synthase sequence illustrated in Figure 4.6, whereas oligonucleotide PfCS(SEQ3) was identical to nucleotides 716-732 (inclusive) of the sense strand. The oligonucleotides were synthesised and purified as described in Section 2.2.9.

A 10 µl volume of the circularised genomic DNA preparation was added to 50 pmoles of each primer and 75 nmoles of each deoxynucleoside 5' triphosphate (dATP, dCTP, dGTP, dTTP) in a 0.1 ml volume of 1x PCR reaction buffer (supplied by the manufacturer), supplemented with BSA to a concentration of 0.1 mg/ml. The reaction was heated to 94°C for 10 min to denature the DNA fully, before 2.5 U *Taq* DNA polymerase were added and the amplification programme started. The amplification programme is described in Section 4.2.3. The amplification reaction was repeated with control samples in which either the target DNA, PfCS(SEQ2) or PfCS(SEQ3) was replaced with water.

5.2.4 Cloning and sequencing of the 1.6 kb amplification product

The 1.6 kb amplification product was treated with *E. coli* DNA polymerase I to ensure it possessed blunt-ended termini and ligated into pUC18-*Sma*I/BAP, as detailed in Sections 4.2.4 and 4.2.5. Competent *E. coli* XL1 Blue were then transformed with the ligation reaction and the resulting colonies screened for the presence of a 1.6 kb insert as described in Section 4.2.6. Five recombinants that contained the 1.6 kb insert, derived from duplicate amplification reactions (INV1 and INV2), were selected for nucleotide sequencing. These were designated PfCS(INV1-5), PfCS(INV2-2), PfCS(INV2-8), PfCS(INV2-9) and PfCS(INV2-13).

The maxi-prep method, (Section 2.2.16) was used to obtain a large scale preparation of plasmid DNA; regions of the insert were then sequenced by the dideoxy-sequencing method of Sanger *et al.* [1977] (Section 2.2.17). Figure 5.2 illustrates the regions of the amplification product sequenced in each clone. Additional 17-mer oligonucleotides, designed to previously determined sequence, were used as sequencing primers. The nucleotide sequence was then translated into amino acid sequence using STADEN PLUS Version 6 (Amersham International plc.).

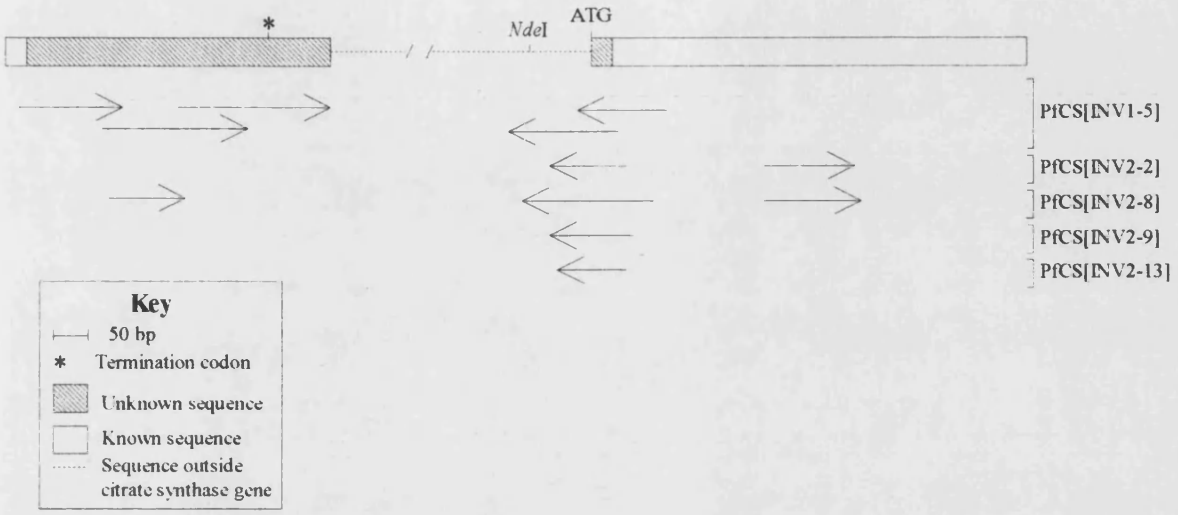


Figure 5.2 Sequencing strategy for the determination of the remainder of the *P. furiosus* citrate synthase gene, amplified from genomic DNA by inverse PCR. The unshaded bar represents known sequence whereas the shaded bar represents desired sequence. The dashed line represents nucleotide sequence outside the citrate synthase coding region and includes the *Nde*I site at which the genomic fragment was circularised prior to the inverse PCR reaction. The arrows represent the sequence obtained by each sequencing primer from its plasmid target which is indicated in the right-hand margin.

5.3 RESULTS

5.3.1 Southern hybridisation analysis of *P. furiosus* genomic DNA

Oligonucleotides PfCS(N) and PfCS(P) bound specifically to fragmented DNA at 55°C, as shown by the single hybridisation signal in each sample lane. Figures 5.3, 5.4 and 5.5 illustrate the binding of PfCS(N) to various digests of *P. furiosus* genomic DNA. As can be seen from Figures 5.3-5.5, digestion of *P. furiosus* DNA with a number of commonly-used restriction endonucleases revealed that in most samples, PfCS(N) hybridised to a gene fragment of 4.5 kb or more.

Oligonucleotide PfCS(N) did not hybridise convincingly to genomic DNA digested with *Xba*I or *Xba*I+*Bam*HI. This suggested that the amino-terminal region of the gene, which is complementary to PfCS(N), contained the recognition site for *Xba*I. This proposal was supported by the specific hybridisation of PfCS(P) to genomic DNA digested with *Xba*I+*Eco*RI (Figure 5.6). The size of the hybridising fragment was indistinguishable from that produced by digestion with *Eco*RI alone. As the 760 bp gene fragment, sequenced in the preceding chapter, did not contain an *Eco*RI restriction site, it was deduced that the *Eco*RI and *Xba*I sites were in close proximity, with the *Eco*RI site located upstream of the coding region of the gene. Thus, a screen of double digests, featuring *Eco*RI, was used to locate a fragment of suitable size for the formation of a genomic library. This series of digests revealed that the region of the gene complementary to PfCS(P) was located on fragments of 4.5 kb or more (Figure 5.6) except when genomic DNA was digested with *Eco*RI+*Hin*dIII. The 1.6 kb fragment generated by this digest was selected for the formation of a size-restricted genomic library as it possessed non-identical termini.

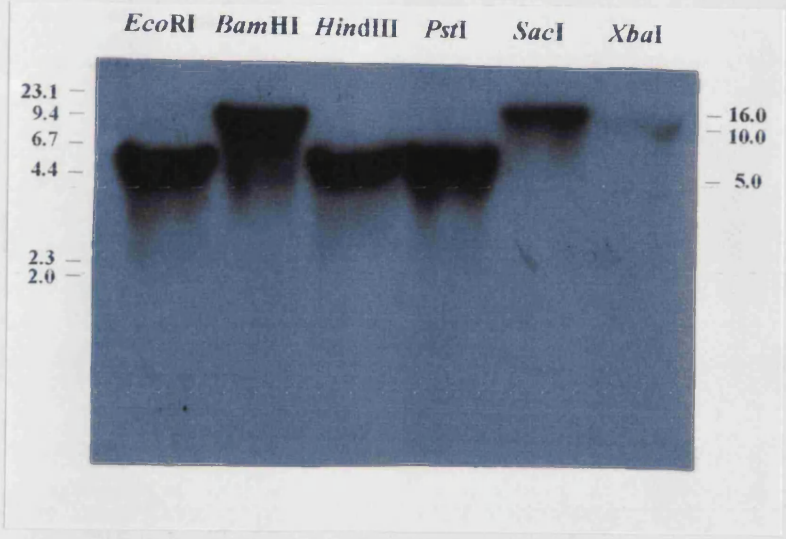


Figure 5.3 Southern hybridisation analysis of *P. furiosus* genomic DNA restricted with *Eco*RI, *Bam*HI, *Hin*dIII, *Pst*I, *Sac*I and *Xba*I using radiolabelled PfCS(N) as a probe. The position of DNA size markers, consisting of *Hin*dIII-digested λ DNA, are indicated in the left hand margin. The approximate sizes of the hybridising fragments (in kb) are indicated in the right hand margin.

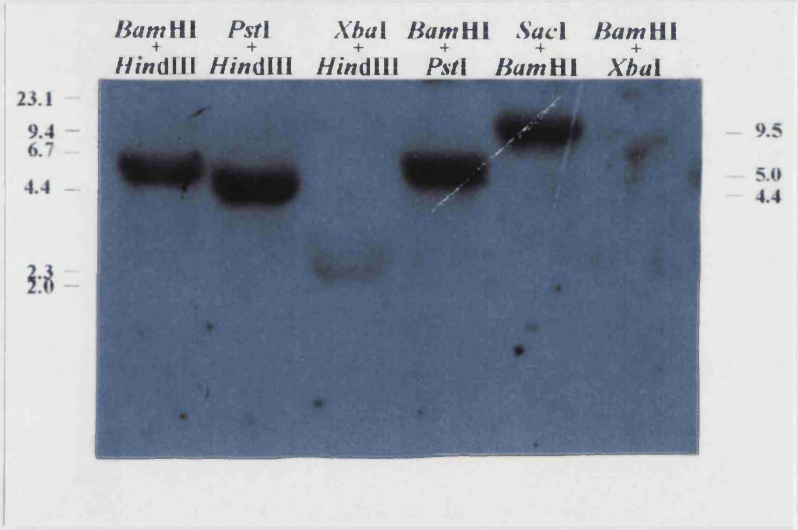


Figure 5.4 Southern hybridisation analysis of *P. furiosus* genomic DNA restricted with *Bam*HI+*Hin*dIII, *Pst*I+*Hin*dIII, *Xba*I+*Hin*dIII, *Bam*HI+*Pst*I, *Bam*HI+*Sac*I and *Bam*HI+*Xba*I using radiolabelled PfCS(N) as a probe. The position of DNA size markers, consisting of *Hin*dIII-digested λ DNA, are indicated in the left hand margin. The approximate sizes of the hybridising fragments (in kb) are indicated in the right hand margin.

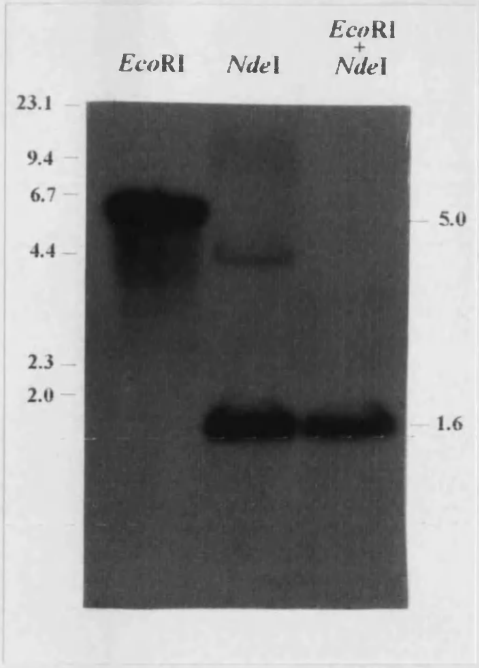


Figure 5.5 Southern hybridisation analysis of *P. furiosus* genomic DNA restricted with *Eco*RI, *Nde*I and *Eco*RI+*Nde*I using radiolabelled PfCS(N) as a probe. The position of DNA size markers, consisting of *Hin*dIII-digested λ DNA, are indicated in the left hand margin. The approximate sizes (in kb) of the hybridising fragments are indicated in the right hand margin.

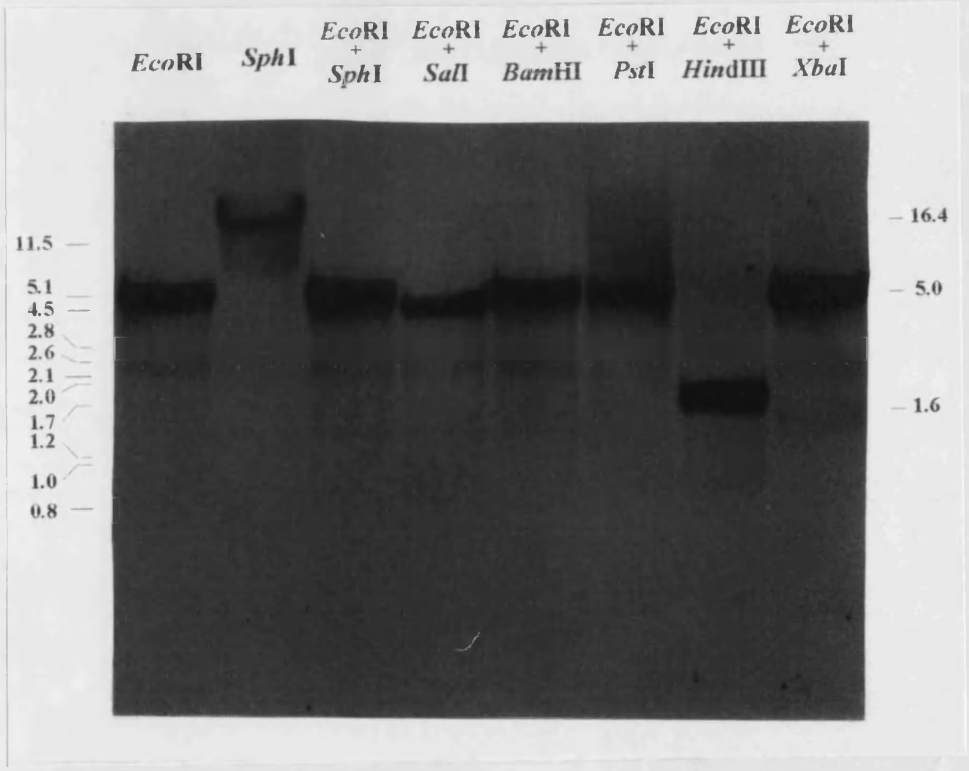


Figure 5.6 Southern hybridisation analysis of *P. furiosus* genomic DNA restricted with *Eco*RI, *Sph*I, *Eco*RI+*Sph*I, *Eco*RI+*Sal*I, *Eco*RI+*Bam*HI, *Eco*RI+*Pst*I, *Eco*RI+*Hin*dIII and *Eco*RI+*Xba*I using Digoxygenin-labelled PfCS(P) as a probe. The position of DNA size markers, consisting of *Pst*I-digested λ DNA, are indicated in the left hand margin. The approximate sizes (in kb) of the hybridising fragments are indicated in the right hand margin.

5.3.2 Attempted formation of a size-restricted genomic library

A 1.6 kb fragment, produced by digestion of genomic DNA with *Eco*RI+*Hin*dIII is presented in Figure 5.6. Although Southern hybridisation analysis with labelled PfCS(P) was not performed on *Hin*dIII-digested genomic DNA, the following evidence inferred that the fragment was doubly-cut. First, as detailed in Section 5.3.1, the position of an *Eco*RI site was mapped to the start of the citrate synthase gene. Second, no *Hin*dIII or *Eco*RI sites were observed in the 760 bp fragment amplified, cloned and sequenced in the preceding chapter. Finally, PfCS(N) hybridised to a 5 kb fragment of *Hin*dIII-digested genomic DNA. Figure 5.7 illustrates the mapping of these sites and demonstrates the location of the coding region within the 1.6 kb, doubly-cut fragment.

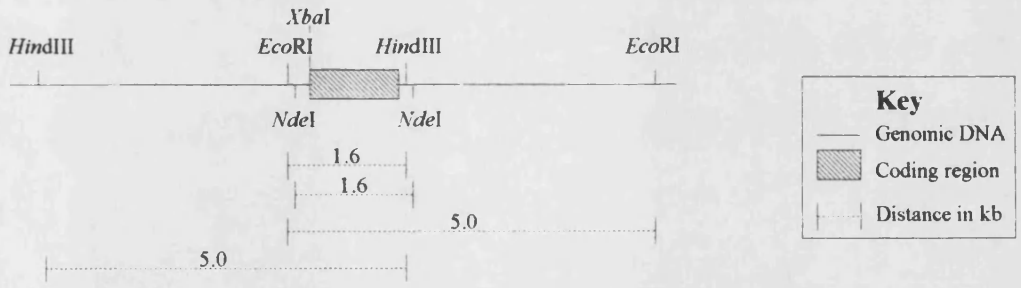


Figure 5.7 Mapping of the *Eco*RI, *Hin*dIII, *Xba*I and *Nde*I restriction sites surrounding the coding region of the *P. furiosus* citrate synthase gene

In parallel with the PCR approach, several attempts were made to clone the 1.6 kb hybridising fragment, produced by digestion of genomic DNA with *Eco*RI and *Hin*dIII, by the construction of a size-restricted genomic library. However, the protocol described in Section 5.2.2 resulted in a library of limited size (less than 100 colonies) and subsequent screening did not reveal the presence of a plasmid that hybridised to either oligonucleotide PfCS(N) or PfCS(P). In light of the success of the inverse PCR procedure in amplifying the remainder of the citrate synthase gene, this cloning procedure was abandoned.

5.3.3 Amplification of the remainder of the citrate synthase gene by inverse PCR

Southern hybridisation analysis of *Nde*I-restricted, *P. furiosus* DNA revealed that oligonucleotide PfCS(N) hybridised specifically to a 1.6 kb fragment (Figure 5.5). As the hybridising fragment possessed cohesive termini, and was under 2.0 kb in size, it proved an ideal target for an inverse PCR amplification. The amplification reaction, described in Section 5.2.3, resulted in the observation of a major DNA product when analysed by agarose gel electrophoresis, as illustrated in Figure 5.8.

5.3.4 Cloning and sequencing of the 1.6 kb amplification product

The amplification product was successfully ligated into pUC18-SmaI/BAP and transformed into *E. coli* XL1 Blue, as detailed in Sections 5.2.4. From two distinct amplification reactions (INV1 and INV2), five recombinants were selected for nucleotide sequencing; these were designated PfCS(INV1-5), PfCS(INV2-2), PfCS(INV2-8), PfCS(INV2-9) and PfCS(INV2-13).

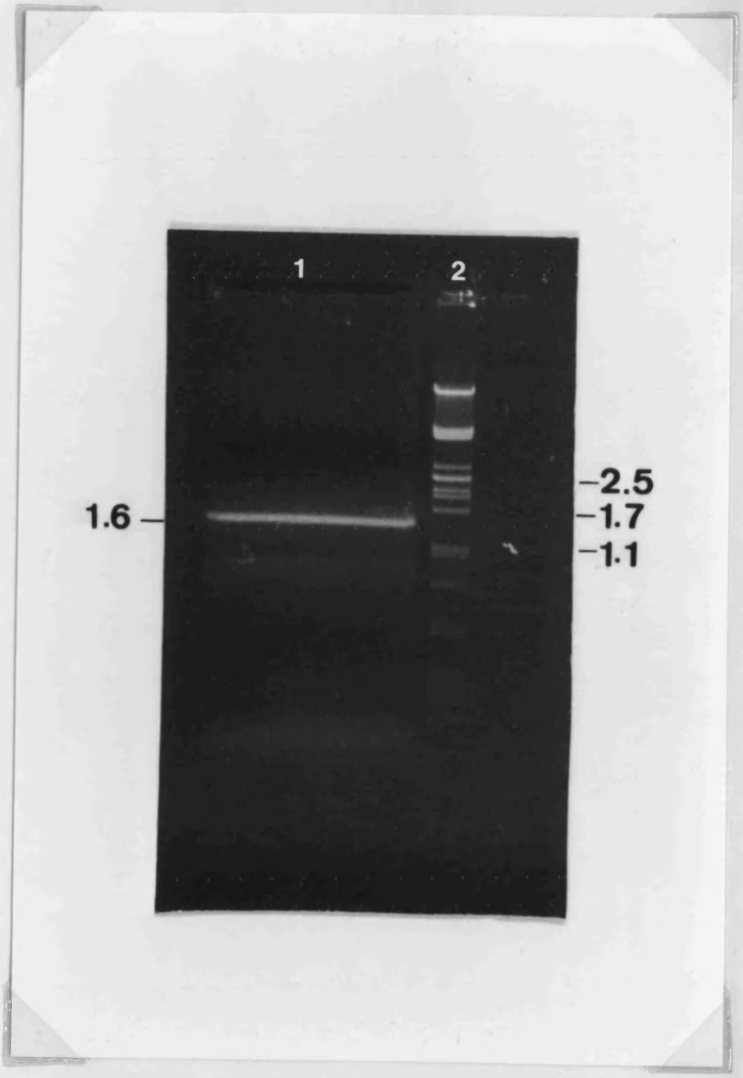


Figure 5.8 Preparative agarose gel of a 1.6 kb fragment of the *P. furiosus* citrate synthase gene, amplified from genomic DNA by inverse PCR.

Lane1: Amplification product

Lane 2: DNA size markers consisting of 0.5 μg *Hin*dIII-digested, λ DNA.

The length (in kb) of selected DNA size markers is indicated in the right hand margin.

The region of the insert encoding the amino-terminus of the gene was sequenced from all five recombinant plasmids, in order to verify the unexpected position of the start codon (Figure 5.9). As mentioned in Section 3.4, amino-terminal analysis of the purified gene product resulted in a major and minor amino acid sequence (Figures 3.3 and 3.5). Analysis of the gene sequence confirmed that the protein had been cleaved during purification and that the minor sequence constituted a faint signal from the full length gene product. In four recombinant plasmids, the Met start codon was located 27 bp upstream of the codon encoding the first residue of the major (cleaved) amino acid sequence. The insert of PfCS(INV2-8) did not possess an in-frame Met codon at this position, possessing instead, an ACG, Thr codon. This is likely to result from an error in the amplification reaction. *Taq* DNA polymerase errors and the potential problems of using a PCR-mediated approach to clone a gene are discussed in Section 5.4.3.

The region of the insert encoding the carboxyl-terminus was sequenced from recombinants PfCS(INV1-5) and PfCS(INV2-8) (Figure 5.10). PfCS(INV1-5) contained a frame shift mutation due to a deletion at position 136 in Figure 5.10; the position of the deleted nucleotide was confirmed by the sequence of PfCS(INV2-8). The translated sequence exhibited 52 % identity to the sequence of the *T. acidophilum* citrate synthase as shown in Figure 5.11, thus confirming that the expected region had been amplified.

```
catatgtccacaaatgttgaacattgtcccgatatttttgcaatatcattataaataata  60

                                N  T  E  K  Y  L  A  K  G  L  E  11
tctctacgaaacattttcagaggtgATGAATACGGAAAAATACCTTGCTAAAGGTCTAGA 120

  D  V  Y  I  D  Q  T  N  I  C  Y  I  D  G  K  E    27
AGATGTCTACATAGATCAAACAAATATATGCTATATCGATGGAAAAGAG  169
```

Figure 5.9 Nucleotide sequence across the amino-terminus of the *P. furiosus* citrate synthase gene. Non-coding sequence is shown in lower case.

```
   M  G  A  G  H  R  V  Y  K  T  Y  D  P  R  A  R  I  F  K    19
TAATGGGTGCTGGTCATAGGGTGTATAAAACCTATGATCCAAGGGCTAGAATATTTAAAA  60

K  Y  A  S  K  L  G  D  K  K  L  F  E  I  A  E  R  L  E  R    39
AATATGCCTCCAAATTAGGAGATAAAAAGCTCTTCGAGATAGCTGAAAGATTAGAGAGGT 120

L  V  E  E  Y  L  S  K  K  G  I  S  I  N  V  D  Y  W  S  G    59
TAGTAGAAGAATACCTAAGCAAAAAAGGAATTAGTATAAATGTTGACTACTGGTCAGGAT 180

L  V  F  Y  G  M  K  I  P  I  E  L  Y  T  T  I  F  A  M  G    79
TAGTTTTCTATGGAATGAAGATTCCAATAGAGCTTTACACAACAATATTTGCAATGGGTA 220

R  I  A  G  W  T  A  H  L  A  E  Y  V  S  H  N  R  I  I  R    99
GAATTGCGGGATGGACTGCTCATTTAGCTGAGTATGTTTCTCACAATAGAATTATTAGAC 280

P  R  L  Q  Y  V  G  E  I  G  K  K  Y  L  P  I  E  L  R  R   119
CCAGGTTGCAGTATGTAGGAGAAATAGGAAAGAAGTACCTACCCATAGAATTAAGGAGGT 340

*
AGataaaatgagcattagacttccaccaagaaggaaaagttattgaagttgtcaaccggc 400

tctctaaagaccctaatgagccca  424
```

Figure 5.10 Nucleotide sequence of the carboxyl-terminal region of the *P. furiosus* citrate synthase gene. The translation termination signal is represented by an asterix; non-coding sequence is shown in lower case.

```
P. furiosus          1 MGAGHRVYKTYDPRARIFKKYASKLGDK.....KLFEIAERLERLVEEYL 45
                       || |||||||||||| ||| | || |     | ||| ||
T. acidophilum     258 MGFGHRVYKTYDPRAKIFKGIAEKLSSKKPEVHKVYEIATKLEDFGIKAF 307

P. furiosus         46 SKKGISINVDYWSGLVFYGMKIPI..ELYTTIFAMGRIAGWTAHLAEYV. 92
                         |||  | || || |      |     ||  ||  |  || ||  |||
T. acidophilum     308 GSKGIYPNTDYFSGIVYMSIGFPLRNNIYTALFALSRVTGWQAHFIEYVE 357

P. furiosus         93 SHNRIIRPRLQYVGEIGKKYLPIELRR 119
                         | ||||  |||    || ||  |
T. acidophilum     358 EQQRLIRPRAVYVGPAERKYVPIAERK 384
```

Figure 5.11 Alignment of the amino acid sequence encoded by part of the 1.6 kb amplification product with primary sequence in the region of the carboxyl-terminus of the *T. acidophilum* citrate synthase [Sutherland *et al.*, 1990]. The alignment was obtained using the BESTFIT program of GCG Sequence Analysis Software (University of Wisconsin). Identical residues are indicated by |.

5.4 DISCUSSION

5.4.1 The cloning strategy - Southern hybridisation

Southern hybridisation analysis pinpointed fragments of *P. furiosus* genomic DNA that were complementary to the labelled oligonucleotide probe. The specificity of oligonucleotides PfCS(N) and PfCS(P) for the *P. furiosus* citrate synthase gene was demonstrated by the single, major hybridisation signal in each sample lane of the Southern blots (Figures 5.3-5.6). This was expected as the oligonucleotides, when used as PCR primers, promoted the amplification of a 760 bp fragment of a putative, citrate synthase gene (Sections 4.3.1 and 4.4.1). Figure 5.5 shows a minor, hybridising band at 4.4 kb in the *Nde*I-digested, genomic DNA sample lane. This is due to incomplete digestion of genomic DNA by *Nde*I; according to the manufacturer, the enzyme is less efficient and less stable than commonly-used restriction endonucleases.

Unfortunately, all attempts to construct a genomic library of the 1.6 kb fragments were unsuccessful. There are a number of reasons why this approach was problematic, such as the efficiency of the T4 DNA ligase, the amount and quality of the vector and insert DNA, the ratio of the vector to insert in the ligation reaction and the efficiency of transformation of competent *E. coli* XL1 Blue. The quality of the insert DNA could have been improved considerably, simply by digesting genomic DNA with *Eco*RI and then purifying (from an agarose gel) the 5.0 kb fragments that are known to hybridise to PfCS(N) (Figure 5.3). These fragments alone, could then be digested with *Hin*dIII and the desired 1.6 kb fragments extracted from an agarose gel. By this, the efficiency of the ligation would be improved by the removal of the majority of the 1.6 kb singly-cut fragments.

The transformation efficiency may also have affected the outcome of this approach. There are numerous methods for transforming *E. coli*. The method chosen here [Hanahan, 1985] resulted in transformation efficiencies of 3-6 x 10^7 transformants/ml cells/μg pUC18, which should have been adequate for the construction of a size-restricted genomic library. Higher transformation efficiencies, in the range of 10^9-10^{10} transformants/ml cells/μg pUC18, can be obtained either by electroporation [Dower *et al.*, 1988], by including either dimethyl sulphoxide (DMSO) or dithiothreitol (DTT) in the transformation buffer [Hanahan, 1983] or by using commercially-available, competent cells. However, in light of the success of the inverse PCR amplification, these approaches were not pursued further.

5.4.2 The cloning strategy - inverse PCR

The PCR approach was applied to the cloning of the *P. furiosus* citrate synthase gene as the formation of a gene library was proving difficult. The oligonucleotide PfCS(N) hybridised to a 1.6 kb, *Nde*I-digested fragment of *P. furiosus* genomic DNA, as presented in Figure 5.5. Mapping of the gene (Figure 5.7) suggested that the entire coding region was contained within this fragment which, by virtue of its cohesive

termini, was a successful target in an inverse PCR amplification (Figure 5.1, Section 5.2.3). By this, a 1.6 kb fragment, containing the regions of nucleotide sequences flanking the sequence determined in the pre-ceding chapter was amplified. Following the ligation of this product into pUC18-*Sma*I/BAP and the transformation of *E. coli* XL1 Blue with the recombinant plasmid, the undetermined regions of the gene were sequenced.

The *Nde*I site, upon which the 1.6 kb fragments were circularised, prior to the inverse PCR amplification, was located 85 bp upstream of the translation start codon, as shown in Figure 5.9. This demonstrated that the mapping was correct and the entire coding region was contained within the 1.6 kb, *Nde*I-digested genomic fragment. As mentioned previously (Section 5.3.4 and Section 3.4), the translation start codon, ATG, was located 27 bp upstream of the codon for Gly9, the residue that was inferred by protein microsequencing to be the amino-terminal residue. By using inverse PCR, it was possible to pinpoint this discrepancy, as well as confirming the nucleotide sequence at the amino-terminus which was previously obscured by the use of redundant PCR primers (Section 4.4.1).

The translated nucleotide sequence of the portion of the *P. furiosus* citrate synthase gene that encoded the carboxyl-terminus of the enzyme was aligned to the primary sequence of the *T. acidophilum* enzyme (Figure 5.11). The two sequences exhibited 52 % identity across this region, thus confirming that the inverse PCR amplification reaction had completed the determination of the gene sequence. A complete analysis of the total gene sequence, including a multiple sequence alignment of the primary sequence of all known citrate synthases, comprises Chapter 7.

5.4.3 PCR-mediated gene cloning - the problem of DNA polymerase fidelity

Prior to reporting the expression of the *P. furiosus* citrate synthase gene in *E. coli* JM105 (Chapter 6) and to analysing the gene sequence (Chapter 7), it is appropriate here, at the completion of the sequence-determination, to discuss the potential problems of polymerase errors. Providing oligonucleotide primers are specific for the gene of interest, PCR is a quick and effective method for obtaining nucleotide sequence, especially, as in this case, when the construction of a genomic library was not successful.

Figure 5.12 illustrates the regions of the *P. furiosus* gene that were sequenced from subclones derived from distinct amplification reactions. This diagram includes the nucleotide sequencing of an additional amplification product, subclone PfCS[TOT5], described in Chapter 6 (Section 6.2.3), which encompasses the entire coding region of the *P. furiosus* citrate synthase. On sequencing these regions, two polymerase errors were observed, a frame shift mutation due to a nucleotide deletion in subclone PfCS(INV1-5) and a nucleotide substitution in subclone PfCS(INV2-8) (Section 5.3.4). The sequencing of repetitive subclones from independent amplification reactions confirmed the correct nucleotide sequence.

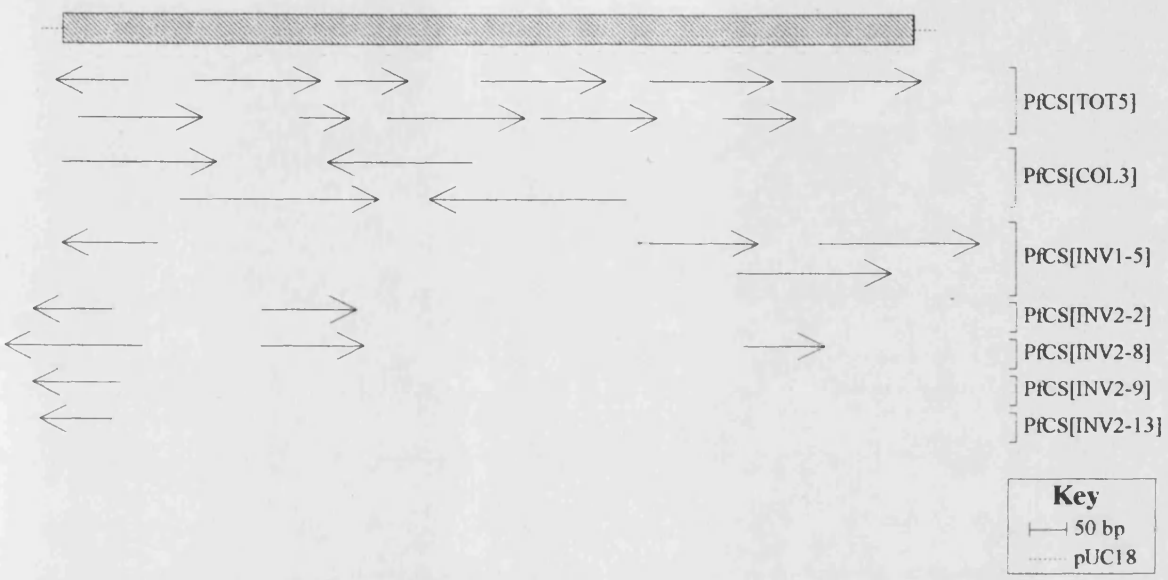


Figure 5.12 Regions the *P. furiosus* citrate synthase gene sequenced from subclones of induvidual amplification reactions to confirm the gene sequence. The hatched bar represents the gene coding region; the arrows represent the sequence obtained by each sequencing primer from its plasmid target, indicated in the right-hand margin. Subclone PfCS[TOT5] is described in Chapter 6.

The fidelity of DNA polymerases has been studied extensively to provide both an insight into cellular mutation rates as well as an appraisal of techniques involving *in vitro* enzymatic DNA synthesis [Bloch, (1991) and references therein]. The most obvious way of assessing the fidelity of *Taq* DNA polymerase is to sequence several subclones derived from a single, amplification reaction. However, this is both labour-intensive and produces a statistically small sample. Two alternative methods have been developed to gauge the fidelity of DNA polymerases used in amplification reactions. The method of Tindall and Kunkel [1988] exploits the *lacZα* gene in the cosmid phage M13mp2. When propagated on media containing X-gal and IPTG, the phage produces blue plaques if the *lacZα* gene is functional and colourless plaques if it is inactive. By using a thermostable DNA polymerase to 'fill in' the nucleotide sequence in a gapped M13mp3 target, statistical data can be generated by counting the resultant blue or white plaques. A variation of this approach assesses fidelity by the rate of reversion of a translation termination codon, engineered into the M13mp3 *lacZα*, to allow the translation of the full-length and active gene product.

The second method [Keohavong and Thilly, 1989; Keovahong *et al.*, 1993] uses radiolabelled primers to amplify a sequence of DNA. The double-stranded product is then heat-denatured and allowed to reanneal to produce a mixture of homoduplexes and heteroduplexes consisting of two wild-type strands and a wild-type and a mutant strand, respectively. The excess of wild-type product should minimise the formation of mutant homoduplexes. When analysed by denaturing gradient gel electrophoresis the heteroduplexes, which are less stable due to the base-mismatch, unwind at a lower concentration of denaturant and are thus retarded in the gel. As the primers are radiolabelled, the amount of radioactivity in the homo- and heteroduplexes can be measured thereby giving an assessment of fidelity. Additionally, by varying the numbers of cycles in the assessed amplification, mutations that are preferentially amplified can be observed as dense, heteroduplex bands in the denaturing gradient gel.

It is difficult to amalgamate the different studies due to the variation in reaction conditions and in the expression of fidelity. Values obtained for *Taq* DNA polymerase range from 2×10^{-4} - 2×10^{-5} errors/bp/duplication, depending on variables such as the template, nucleotide concentration, reaction buffer, Mg^{2+} concentration and annealing temperature - as well as the method of assessment of fidelity [Bloch, 1991]. However, using one method, different species of polymerase can be compared. For example, using the method of denaturing gradient gel electrophoresis [Keohavong *et al.*, 1993] the fidelity of $Vent_R$ DNA polymerase was shown to be over three times higher than for *Taq* DNA polymerase (8.9×10^{-5} and 2.4×10^{-5} errors/bp/duplication, respectively).

The types of polymerase error can be assessed by sequencing the DNA from a *lacZ*α^- plaque or from an mutant strand in a heteroduplex. Base substitutions of the A/T $\rightarrow$ G/C nature are the predominant mutation; base deletions are less common and insertions rarer still. On resolving the substitution event in subclone PfCS(INV2-8), an A/T $\rightarrow$ G/C substitution had occurred. This erroneous substitution was of the type commonly incorporated by *Taq* DNA polymerase and did not occur in other subclones. In a similar manner, the site of the base deletion was easily resolved on sequencing two other subclones from independent amplification reactions. By the nucleotide sequencing of regions of the *P. furiosus* citrate synthase gene from seven subclones derived from four, separate, amplification reactions, the resolved sequence can be declared with confidence. However, if the sequencing of subclones had not resolved these discrepancies, the PCR product could be sequenced directly without cloning, following asymmetric PCR to produce predominantly single-stranded product [Gyllensten and Erlich, 1988]. On sequencing this, polymerase errors, which occur at a low frequency throughout a region of sequence, will be masked by the excess of wild-type, correct sequence.

5.4.4 Conclusions

The remainder of the *P. furiosus* citrate synthase gene has been determined following the amplification of a 1.6 kb DNA fragment from *P. furiosus* genomic DNA. The disadvantage of using this approach, the incorporation of polymerase errors, was alleviated by sequencing multiple subclones from independent amplification reactions. The following chapter (Chapter 6) provides proof that this sequence encodes a functional, hyperthermostable enzyme. With this knowledge, Chapter 7 consists of a detailed analysis of the gene and protein sequence.

CHAPTER 6

EXPRESSION OF THE *P. FURIOSUS* CITRATE SYNTHASE GENE IN *E. COLI* JM105

6.0 INTRODUCTION

The citrate synthase gene of *P. furiosus* was amplified, cloned and sequenced in two parts (Chapters 4 and 5). The expression of the gene in *E. coli* JM105 was necessary both to confirm that the gene encodes a functional enzyme and to produce sufficient quantities of enzyme for detailed study.

To express the citrate synthase gene, the vector pKK223-3 (Pharmacia), shown in Figure 6.1, was chosen. This expression vehicle contains a vector-encoded promoter, ribosome binding site and termination signals thereby eliminating any difficulties that may have arisen if using the *P. furiosus* nucleotide signals within *E. coli* JM105. (Section 7.2.1 describes the putative *P. furiosus* transcription and translation signals.) Thus, providing the ATG start codon is inserted within 10-15 base pairs of the vector-encoded ribosome binding site, the vector is optimised for expression within *E. coli* JM105. As a final amplification was required to obtain the coding region in a single fragment, the correct orientation of the insert could be ensured by incorporating sites for the appropriate restriction endonucleases in the amplification primers.

This chapter details the amplification, cloning and sequencing of the citrate synthase coding region and the purification and characterisation of the recombinant enzyme.

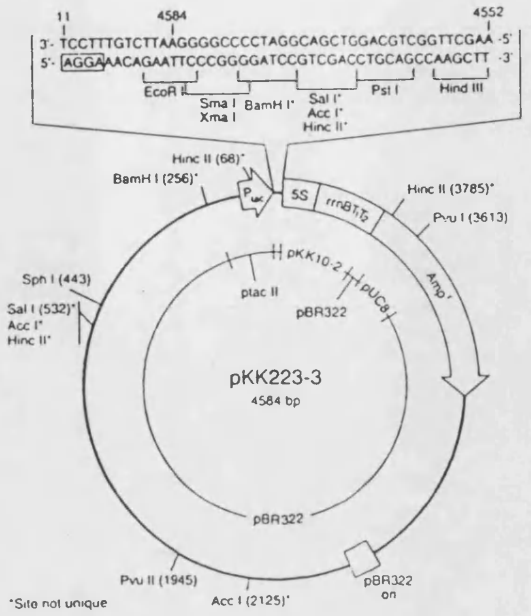


Figure 6.1 The expression vector pKK223-3 (Pharmacia)

6.1 MATERIALS

P. furiosus genomic DNA was generously donated by Dr R. Sharp, Centre for Applied Microbiological Research, Porton Down, UK. Suppliers of restriction enzymes, other DNA-modifying enzymes and radiochemicals are detailed in Section 2.2.1. *E. coli* strains and culture conditions are detailed in Section 2.2.2. The expression vector, pKK223-3, and the pre-packed Superose 12 gel filtration column were obtained from Pharmacia. Centriprep™ concentrators and Matrex™ gel Red A were obtained from Amicon. $Vent_R$ DNA polymerase was obtained from New England Biolabs, Beverly, USA. Glass capillary tubes (1-2 mm in diameter) were supplied by Bilbate Precision Glass, Daventry, UK.

6.2 METHODS

6.2.1 Amplification of the citrate synthase coding region

Oligonucleotide primers were designed to opposite strands of known nucleotide sequence flanking the citrate synthase coding region. The recognition site for endonucleases *Eco*RI and *Hin*dIII was engineered into the forward and reverse primer, respectively. Primer specificity was ensured by including a minimum of nine nucleotides of known sequence in each primer, before and after any base mismatches. The forward primer, designated PfCS(PCR2), was designed to the sense strand of the gene corresponding to nucleotides 72-95 (inclusively) of Figure 5.9. The reverse primer, designated PfCS(PCR3), was designed to the antisense strand of the gene corresponding to nucleotides 348-371 (inclusively) of Figure 5.10. The sequence of each oligonucleotide is illustrated in Figure 6.2. Oligonucleotides were synthesised and purified as described in Section 2.2.9.

The citrate synthase gene was amplified from *P. furiosus* genomic DNA using a high-fidelity $Vent_R$ DNA polymerase [Kong *et al.*, 1993]. As suggested by the manufacturer, the amplification was optimised by assessing different concentrations of $MgSO_4$ in the reaction buffer. A 100 ng quantity of *P. furiosus* genomic DNA was incubated with 50 pmoles of each primer and 30 nmoles of each nucleoside 5' triphosphate in a 0.1 ml volume of 1x PCR reaction buffer (supplied by the manufacturer), supplemented with $MgSO_4$ to a final concentration of 2 mM, 4 mM or 5.5 mM. Following the addition of 1 U $Vent_R$ DNA polymerase, the following programme was repeated 25 times to amplify the coding region: 94°C, 1.25 min; 55°C, 0.5 min; 72°C, 1.5 min. After the final cycle, the reactions were incubated at 72°C for 10 min to ensure primer extension was complete.

A 10 µl volume of each reaction was analysed by agarose gel electrophoresis as described in Section 2.2.6. The amplification was repeated with control samples in which either the target DNA, PfCS(PCR2) or PfCS(PCR3) was replaced with water.

Forward primer

Amino acid	M N T
Nucleotide sequence 5'-3'	AAACATTTTCAGAGGTGATGAATACGGA
PfCS(PCR2) 5'-3'	CATTTTCA**GAATTC**ATGAATACGG

Reverse primer (antisense)

Amino acid	R *
Nucleotide sequence 3'-5'	TCCATCTATTTTACTCGTAATCTGAAGGTGGTTCT
PfCS(PCR3) 3'-5'	TACTCGTAA**TTCGAA**GGTGGTTCT

Figure 6.2 Nucleotide sequence of the forward and reverse primers (PfCS(PCR2) and PfCS(PCR3), respectively) that were used to amplify the *P. furiosus* citrate synthase coding region.

The restriction sites, engineered into the primers, are shown in bold font.

6.2.2 Ligation of the coding region into pKK223-3 and transformation of *E. coli* JM105

A 10 µl volume of the amplification product was purified from the PCR reaction constituents using the Geneclean method (Section 2.2.7) and digested sequentially with *Hin*dIII and *Eco*RI (Section 2.2.5). Following agarose gel electrophoresis to remove the terminal fragments from the full-length coding region, the product was purified again using Geneclean. In a similar manner, a 0.5 µg quantity of pKK223-3 was digested sequentially with *Hin*dIII and *Eco*RI and the intervening fragment removed by agarose gel electrophoresis. The vector was then purified using Geneclean. Equal amounts (~50 ng) of vector and insert were incubated overnight at 15°C in a 20 µl volume of 1x ligase buffer (supplied by the manufacturer) containing 8 U T4 DNA ligase. To assess the degree of vector religation, a control reaction was prepared in which the amplification product was replaced by distilled water.

Competent *E. coli* JM105, prepared by the method of Hanahan [1985], were then transformed with a 2 µl volume of the ligation reaction by the heat shock method (Section 2.2.14). Recombinant colonies were screened for the presence of the insert by isolating the plasmid from each (Section 2.2.15), digesting the preparation with *Hin*dIII and *Eco*RI, and analysing the resultant fragments by agarose gel electrophoresis. Clone PfCS(EXP5) was selected for further analysis.

6.2.3 Determination of the nucleotide sequence of the coding region

To check the fidelity of the final amplification, the nucleotide sequence of the coding region was determined. However, as pKK223-3 is a low copy-number vector, it was necessary to ligate the coding region into pUC18 in order to obtain a large-scale plasmid preparation.

A 2 µg quantity of pUC18 was sequentially digested with *Hin*dIII and *Eco*RI and purified using Geneclean (Section 2.2.7) following agarose gel electrophoresis to remove the intervening fragment. A similar digestion was used to remove the citrate synthase coding region from a small scale preparation of plasmid isolated from PfCS(EXP5); the coding region was then was purified from an agarose gel using Geneclean. Vector and insert were ligated and transformed into *E. coli* XL1 Blue as described above (Section 6.2.2). After screening, recombinant PfCS(TOT5) was selected for sequencing. The 'maxi-prep' method (Section 2.2.16) was used to obtain a large-scale preparation of purified plasmid. The insert was then sequenced by the dideoxy-sequencing method of Sanger *et al.* [1977], as detailed in 2.2.17. Figure 6.3 illustrates the strategy for sequencing the citrate synthase coding region. Additional 17 base oligonucleotides, designed to previously-determined sequence, were used as sequencing primers. Extra oligonucleotides were synthesised and purified as described in Section 2.2.9.

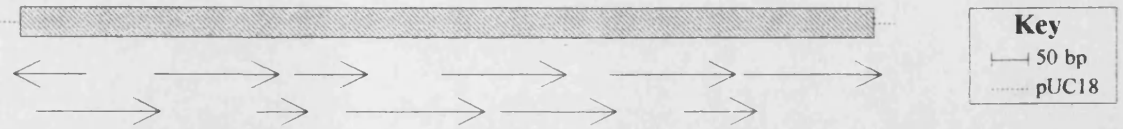


Figure 6.3 Sequencing strategy for the determination of the nucleotide sequence of the *P. furiosus* coding region amplified from genomic DNA. The hatched bar represents the citrate synthase coding region. The arrows represent the nucleotide sequence obtained from each sequencing primer.

6.2.4 Expression and purification of *P. furiosus* citrate synthase from *E. coli* JM105

Recombinant PfCS(EXP5) was cultured overnight, at 37°C, in a 1 l volume of 2 x YT media (distributed between two, 2 l - volume conical flasks) containing ampicillin to a concentration of 100 μg/ml. After 15 h incubation, ampicillin was added to the media, as above, and the culture incubated for a further 2 h. The cells were then harvested by centrifugation at 5,000 x *g* for 20 min. The cell pellet was resuspended in a 20 ml volume of buffer A (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 25 mM NaCl), supplemented with phenylmethanesulfonyl fluoride (PMSF) to a concentration of 1 mM, and disrupted by sonication. Sonication was performed on aliquots of cell suspension of volume 5 ml, using a probe of 15 mm diameter; each 5 ml volume was subjected to sonication of 4 bursts of 20 s at 180 W. Cell debris was then removed by centrifugation at 15,000 x *g* for 20 min. The resultant supernatant was heated to 85°C for 15 min and the denatured proteins of *E. coli* JM105 removed by centrifugation (12,000 x *g*, 20 min).

The heat-treated extract was applied to a Matrex gel Red A column (volume 30 ml), previously equilibrated in buffer A. Flow rates and fraction collecting were controlled by a Gradifrac™ system (Pharmacia). After washing the column with buffer A, the *P. furiosus* citrate synthase was eluted with a 30 ml volume of buffer A containing 5 mM oxaloacetate and 1 mM CoA (pH 8.0 (NaOH)). Fractions were assayed for citrate synthase activity (Section 2.1.2) and the active fractions pooled. The purified enzyme was then concentrated and buffer-exchanged with buffer A using Centriprep concentrators. Enzyme purity was assessed by SDS-PAGE under reducing conditions [Laemmli, 1970] (Section 2.1.4). Protein concentration was measured by the method of Bradford [1976] (Section 2.1.3).

6.2.5 Amino-terminal sequence analysis of the recombinant enzyme

A 20 μg quantity of purified, recombinant citrate synthase was electrophoresed under reducing conditions (Laemmli, 1970) and electrophoretically transferred onto a polyvinylidene difluoride membrane (Immobilon P), as described in Sections 2.1.4 and 2.1.5, respectively. The protein band was cut out, air-dried and subjected to amino-terminal sequence analysis (Section 2.1.6) (J. Young, Zeneca Pharmaceuticals, Macclesfield, UK.).

6.2.6 Determination of the relative molecular mass of the recombinant enzyme

A pre-packed, Superose 12 gel filtration column (Pharmacia) was used to determine the relative molecular mass (M_r) of the recombinant enzyme. Flow-rate control and absorbance measurement at 280 nm were achieved by an FPLC system (Pharmacia). The column was equilibrated in 50 mM potassium phosphate pH 7.0 containing 0.15 M NaCl at a flow-rate of 0.2 ml/min. Each protein sample was then applied separately to the column to eliminate protein aggregation. The excluded volume was determined with Blue Dextran and the included volume with 2,4-DNP-L-lysine. The elution volumes of protein standards, soybean trypsin inhibitor, bovine spleen deoxyribonuclease II, porcine heart malate

dehydrogenase, porcine heart citrate synthase and rabbit glyceraldehyde-3-phosphate dehydrogenase were used to construct a calibration curve. The elution volume of the recombinant, *P. furiosus* citrate synthase was then determined, and its M_r calculated from the calibration data. The M_r of the citrate synthase subunit was determined both experimentally, by SDS-PAGE under reducing conditions (Section 2.1.4) and theoretically, from the translated gene sequence.

6.2.7 Dependence of citrate synthase activity on the concentration of NaCl and KCl

Citrate synthase was assayed in 50 mM EPPS pH 8.0 (HCl) containing 2 mM EDTA as described previously (Section 2.1.2). The enzyme concentration of each assay was 0.25 μg/ml. Activity was assayed in the presence of 0-200 mM KCl or NaCl.

6.2.8 Determination of the kinetic parameters for the recombinant citrate synthase

Citrate synthase was assayed in 50 mM EPPS pH 8.0 (HCl) containing 2 mM EDTA and 0.1 M KCl, as described previously (Section 2.1.2). The dependence of enzyme rate on oxaloacetate was determined using 0.15 μg/ml enzyme and 0.15 mM acetyl-CoA, varying oxaloacetate from 0-100 μM. The dependence of enzyme rate on acetyl-CoA was determined using 0.015 μg/ml enzyme and 0.2 mM oxaloacetate, varying acetyl-CoA from 0-20 μM. Data from quadruplet samples at each substrate concentration were analysed by the direct linear plot [Eisenthal and Cornish-Bowden, 1974] using ENZPACK3 (BIOSOFT, Cambridge, UK.).

6.2.9 Analysis of the thermostability of the recombinant citrate synthase

Pure, recombinant enzyme was diluted into 50 mM potassium phosphate pH 6.5 to a protein concentration of 30 μg/ml. Volumes of diluted enzyme of 50 μl were then drawn into glass capillary tubes (1-2 mm diameter) by gentle suction and the ends sealed in the flame of a Bunsen burner. The sealed tubes were then incubated in a PEG 4000 bath at temperatures of 98°C, 100°C, 102°C and 104°C. At each time point, a tube was removed and cooled on ice for 2 min. After wiping with an acetone-soaked tissue, the tube was broken open to release the contents into a microfuge tube. The preparation was then assayed for residual citrate synthase activity in 50 mM EPPS pH 8.0 (HCl) containing 2 mM EDTA and 0.1 M KCl as described previously (Section 2.1.2). Data was then analysed using QUATTRO PRO software (Borland International, Twyford, UK).

6.3 RESULTS

6.3.1 Amplification and cloning of the *P. furiosus* citrate synthase coding region

Using $Vent_R$ DNA polymerase, a single product of 1.1 kb was amplified, from genomic DNA, using oligonucleotides PfCS(PCR2) and PfCS(PCR3) as primers (Figure 6.4). The product was generated only when the concentration of $MgSO_4$ was increased to 5.5 mM; this is within the concentration range recommended by the manufacturer. The coding region was ligated into pKK223-3, transformed into *E. coli* JM105 and recombinant PfCS(EXP5) selected for analysis. The nucleotide sequence of the plasmid insert of PfCS(EXP5) was determined following the subcloning of the insert into pUC18. The nucleotide sequence did not contain any polymerase-induced errors, as discussed previously (Section 5.4.3). The complete gene sequence is illustrated and analysed in Chapter 7.

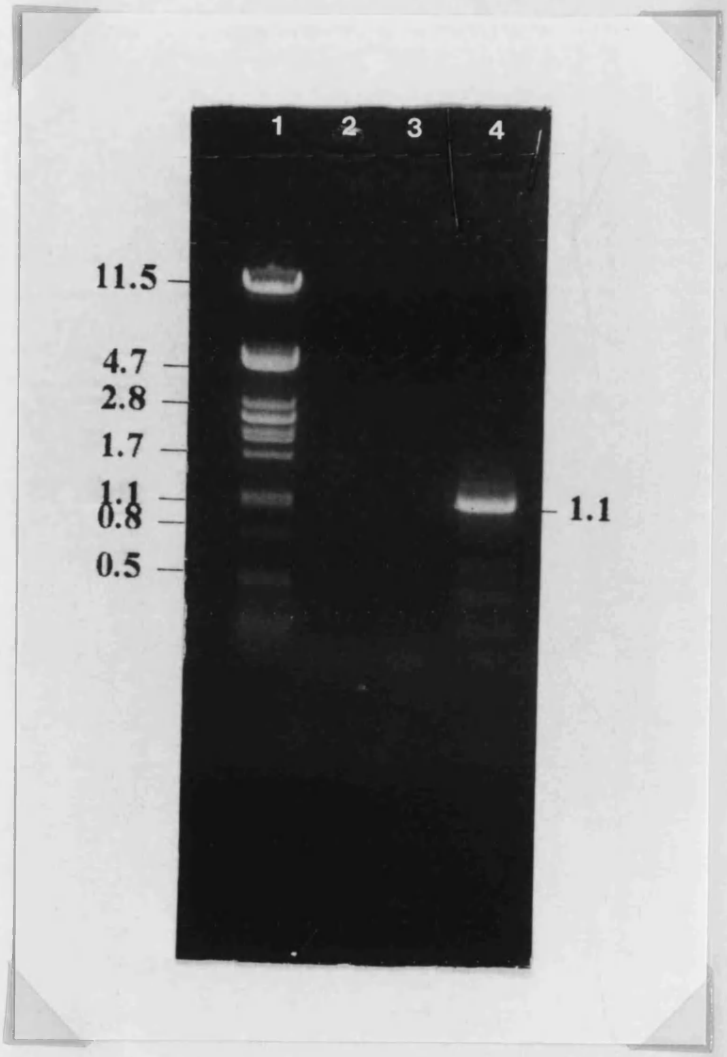


Figure 6.4 Agarose gel illustrating the amplification of the citrate synthase coding region from *P. furiosus* genomic DNA using Vent DNA polymerase

Lane 1: DNA size markers consisting of 0.5μg *Hin*dIII-digested λ DNA
The length in kilobase pairs of selected DNA size markers is indicated in the right hand margin.

Lanes 2-4: Amplification products
The amplification was performed in buffer containing $MgSO_4$ at 2, 4 and 5.5 mM.
Only the reaction in 5.5 mM $MgSO_4$ amplified a 1.1 kb fragment.

6.3.2 Purification of the *P. furiosus* citrate synthase from *E. coli* JM105

Recombinant PfCS(EXP5) expressed the *P. furiosus*, citrate synthase gene constitutively. This was illustrated by comparing an SDS-PAGE analysis of the total cell protein of PfCS(EXP5) with that of *E. coli* JM105 transformed with pKK223-3. As shown in Figure 6.5, the cell extract of PfCS(EXP5) contained an extra protein band at 42.6 ± 2.2 kDa. The recombinant enzyme was purified by an 85°C heat-treatment followed by an affinity dye chromatographic step (Figure 6.6). The purified protein was shown to be homogenous both by Coomassie blue-stained SDS-PAGE and by Superose 12 gel filtration. As a 36-fold purification was required to purify the enzyme to homogeneity (Table 6.1), the recombinant protein constituted 2.8 % of the total *E. coli* JM105 cell protein. Chapter 3 details the partial purification of the enzyme from *P. furiosus* cell extracts whereby a 432.5-fold purification was required to distinguish the citrate synthase protein by SDS-PAGE. Therefore, in the *P. furiosus* cell extracts, the enzyme constituted less than 0.2 % of the total protein.

Step	Volume (ml)	Total enzyme activity (μmol/min)	Total protein (mg)	Specific Activity (μmol/min/mg)	Yield (%)
Cell extract	19	44	53	0.8	100
Heat treatment	15	32	2.9	11	73
Matrex gel Red A	15	27	0.9	30	61

Table 6.1 Purification of *P. furiosus* citrate synthase (recombinant) from *E. coli* JM105 cell supernatant

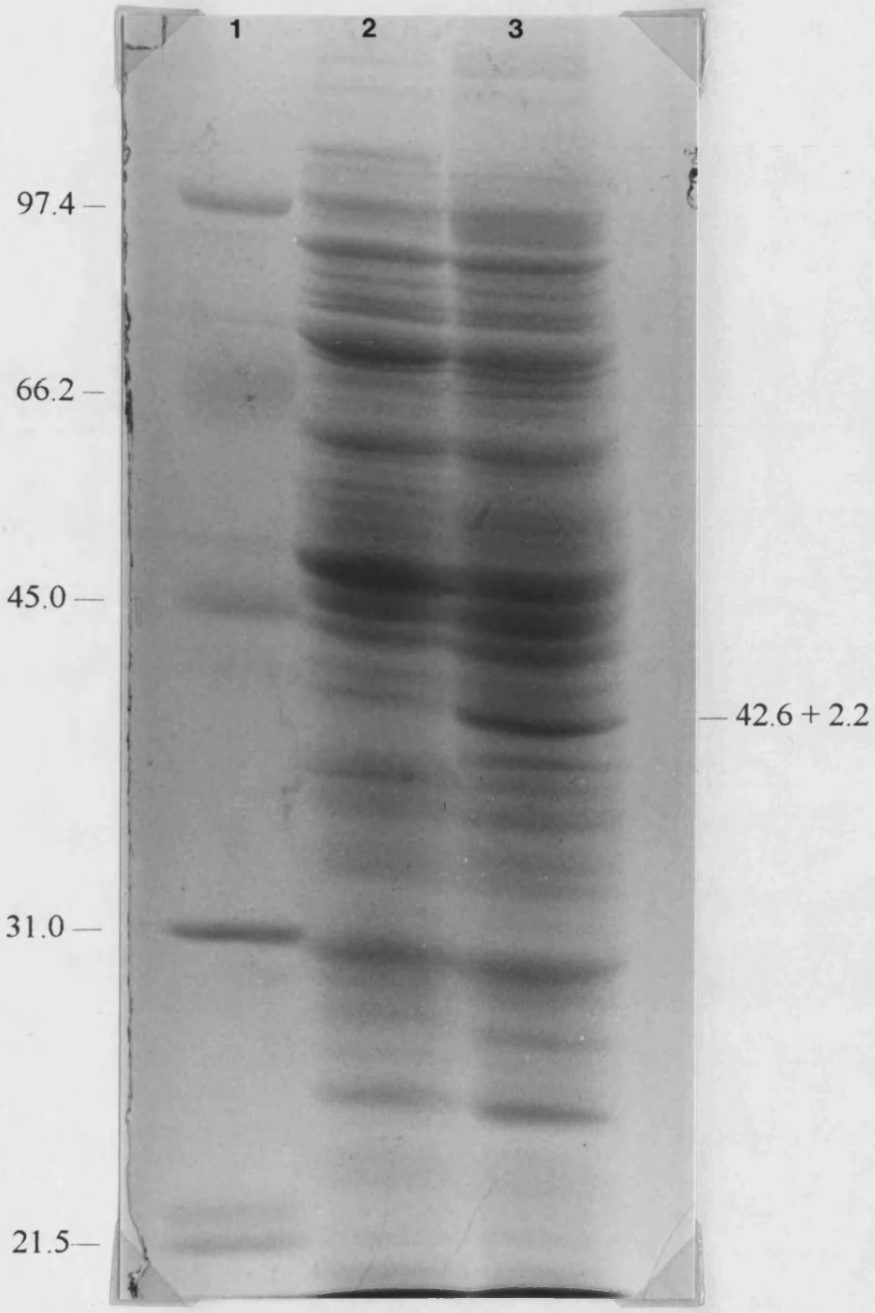


Figure 6.5 SDS-PAGE analysis to illustrate the expression of the *P. furiosus* citrate synthase gene in *E. coli* JM105

Lane 1 Protein standards
The M_r of each protein standard in kDa is indicated in the left margin.
Lane 2 Cell supernatant of *E. coli* JM105 transformed with pKK223-3 (60 μg protein)
Lane 3 Cell supernatant of PfCS(EXP5) (60 μg protein)

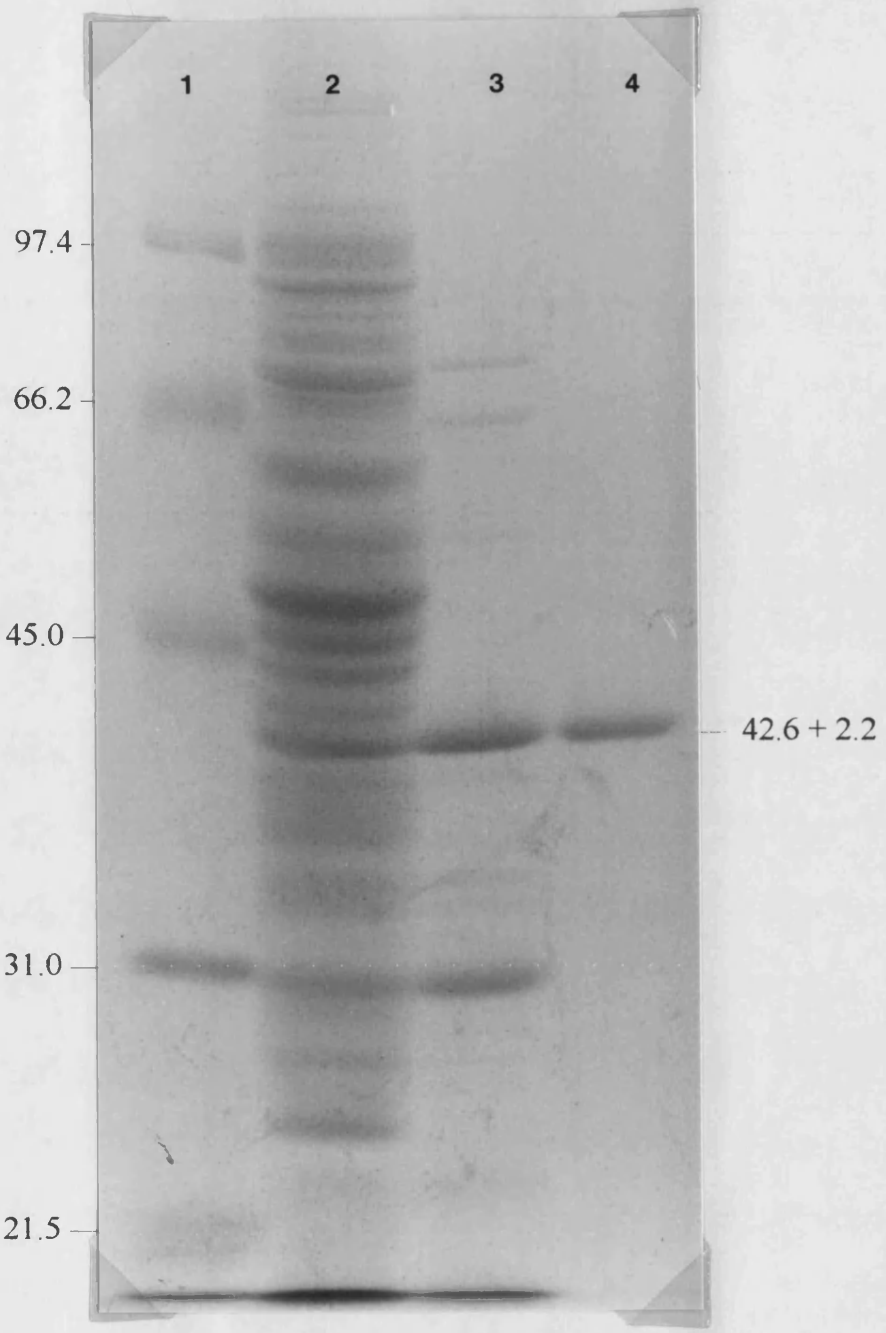


Figure 6.6 SDS-PAGE analysis illustrating the purification of the *P. furiosus* citrate synthase from *E. coli* JM105

Lane 1	Protein standards The M_r of the each protein standard in kDa is indicated in the left margin.
Lane 2	Cell supernatant of PfCS(EXP5) (60 μg protein)
Lane 3	Heat-treated cell supernatant (5 μg protein)
Lane 4	Affinity chromatography of heat-treated cell supernatant (2 μg protein)

6.3.3 Relative molecular mass and oligomeric status of the recombinant protein

A subunit M_r of 40.8 kDa was calculated from the translated nucleotide sequence of the *P. furiosus* citrate synthase coding region. Analysis of the protein by SDS-PAGE agreed with this estimate; a subunit M_r of 42.6 ± 2.2 kDa was determined by comparison to protein standards. The oligomeric status of the enzyme was determined by Superose 12 gel filtration (Section 6.2.6). The distribution coefficient, K_d, of each protein standard was calculated by $K_d = (V_e-V_o)/(V_t-V_o)$ where V_e is the elution volume of the protein standard, V_o is the excluded volume determined with Dextran blue, and V_e is the included volume determined with 2,4-DNP-L-lysine. Figure 6.7 illustrates a calibration curve, constructed from standard proteins, from which the M_r of the *P. furiosus* citrate synthase was determined as 89.7 ± 1.6 kDa. Thus, in common with all archaeal citrate synthases studied to date [Danson, 1988], the *P. furiosus* enzyme is a dimer.

6.3.4 Amino-terminal sequencing of the recombinant enzyme

Microsequencing of the purified, recombinant enzyme showed that translation had been initiated at the expected, ATG codon thereby producing a full length product. However, a major and a minor amino acid signal with each cycle of Edman degradation revealed that ~70 % of the purified enzyme had retained the amino-terminal Met whereas processing within *E. coli* JM105 had removed the Met in ~30 % of the preparation.

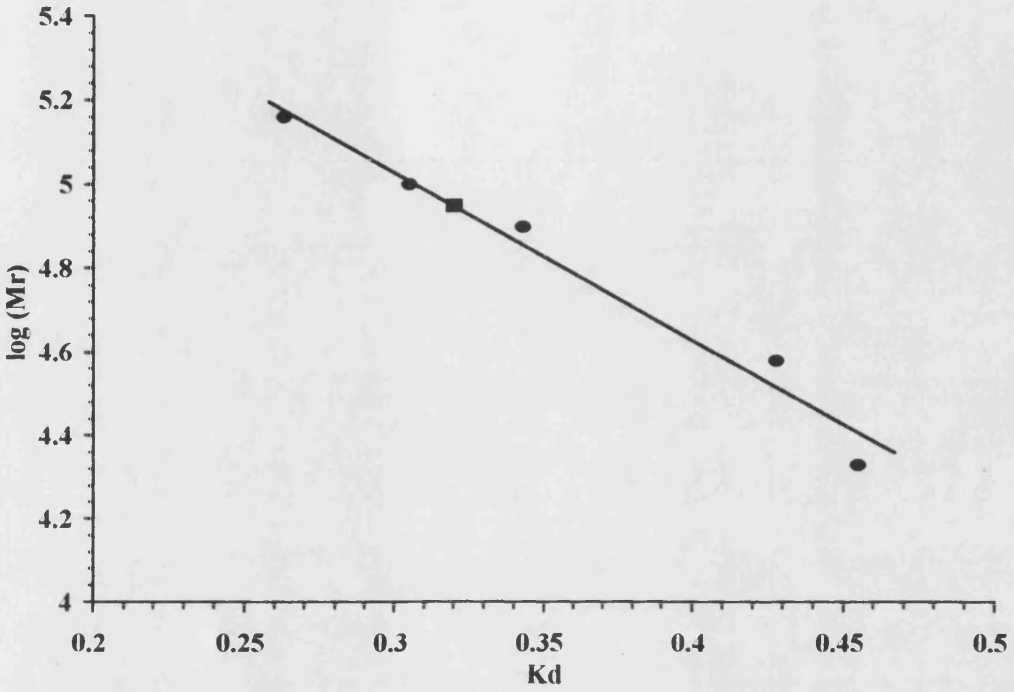


Figure 6.7 Calibration curve of the logarithm of the molecular mass of each standard protein in daltons against the distribution coefficient (K_d) for a Superose 12 gel filtration. A standard curve has been fitted to the data points by linear regression. The ■ represents the data point obtained for the *P. furiosus* citrate synthase.

6.3.5 Enzymatic activity of the recombinant citrate synthase

Prior to the determination of the kinetic parameters of the enzyme, it was necessary to optimise the enzyme assay. As described in Section 2.1.2, the enzyme assay was performed at 55°C which is sub-optimal for the *P. furiosus* enzyme but convenient when considering the thermolability of oxaloacetate ($t_{½}$ (54°C) = 23 min; $t_{½}$ (73°C) = 4 min [M.J. Danson - personal communication]). As the internal environment of *Pyrococcus woesei* was reported to contain concentrations of K^+ of 0.6 M [Scholz *et al.*, 1992], the effect of KCl and NaCl on enzyme activity was measured. Figure 6.8 illustrates the increase in reaction rate (v) of the *P. furiosus* citrate synthase in buffer supplemented with 0 - 200 mM NaCl or KCl, over the rate measured in the absence of NaCl or KCl (v_o). In buffer supplemented with NaCl to a concentration of a 200 mM, the reaction rate increased approximately 5 fold; this was slightly higher than the increase recorded for the assay in 200 mM KCl. In both cases, the rate of increase was non-linear and slowed considerably between 100-200 mM. From these data, the kinetic parameters for the *P. furiosus* citrate synthase were determined at 55°C in 50 mM EPPS pH 8.0 (HCl) containing 2 mM EDTA and 100 mM KCl.

Figures 6.9 and 6.10 illustrate the increase in enzyme rate with increasing concentration of substrates as a Hanes plot of (s/v) against (s), where (s) is the substrate concentration and (v) is the reaction velocity. Analysis of these data by the direct linear plot [Eisenthal and Cornish-Bowden, 1974] resulted in the parameters displayed in Table 6.2. The values for 68 % confidence limits confirm that the data fit the hyperbolic model well. The values for V_{max} show some variation although they are of similar order of magnitude.

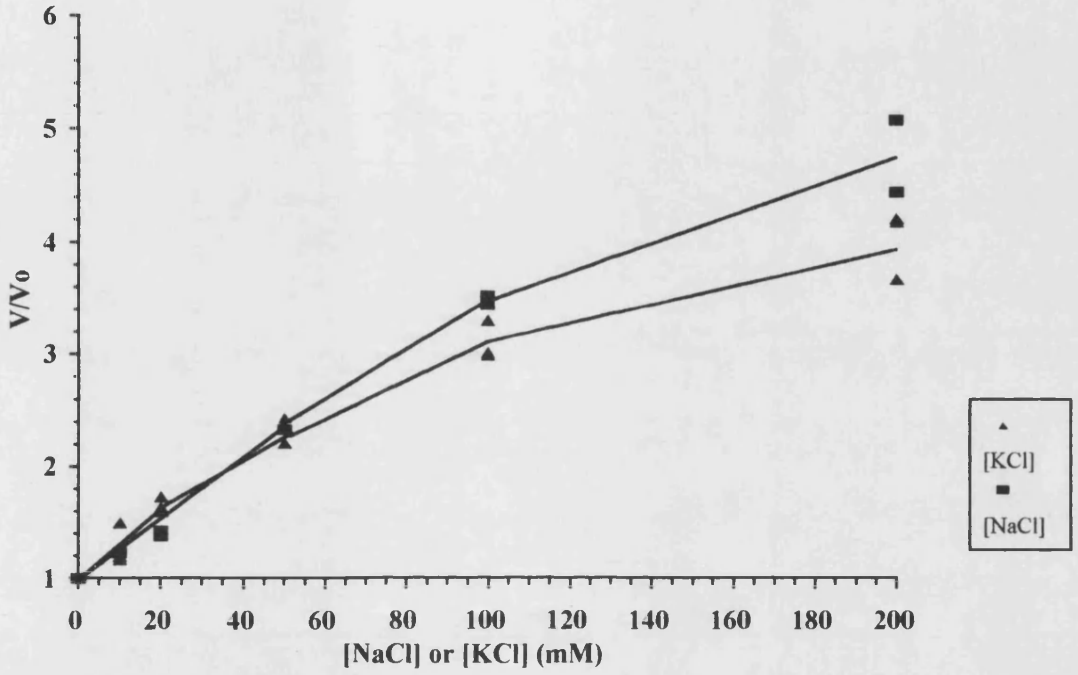


Figure 6.8 Effect of NaCl and KCl on the activity of the *P. furiosus* citrate synthase. The increase in enzyme rate is shown by v/v_o where v is the measured reaction velocity and v_o is the reaction velocity in the absence of added NaCl or KCl.

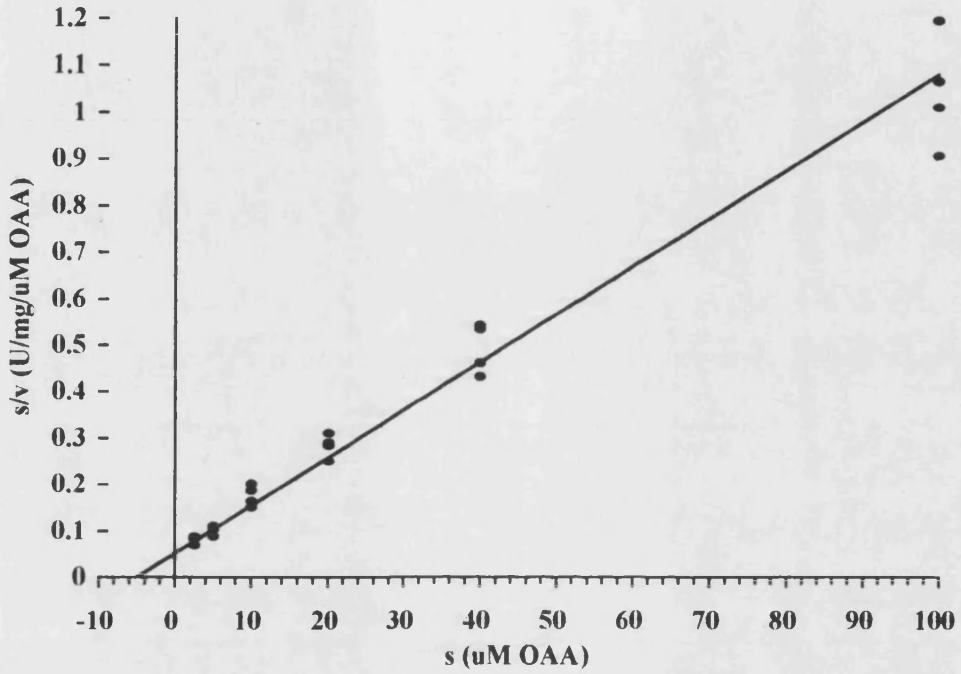


Figure 6.9 Hanes plot of the dependence of reaction rate on the concentration of oxaloacetate for the recombinant *P. furiosus* citrate synthase. Each assay was performed at 55°C and contained 0.15 μg/ml enzyme and 0.15 mM acetyl-CoA. The line is derived from the kinetic parameters obtained from a direct linear plot of the data (Table 6.2).

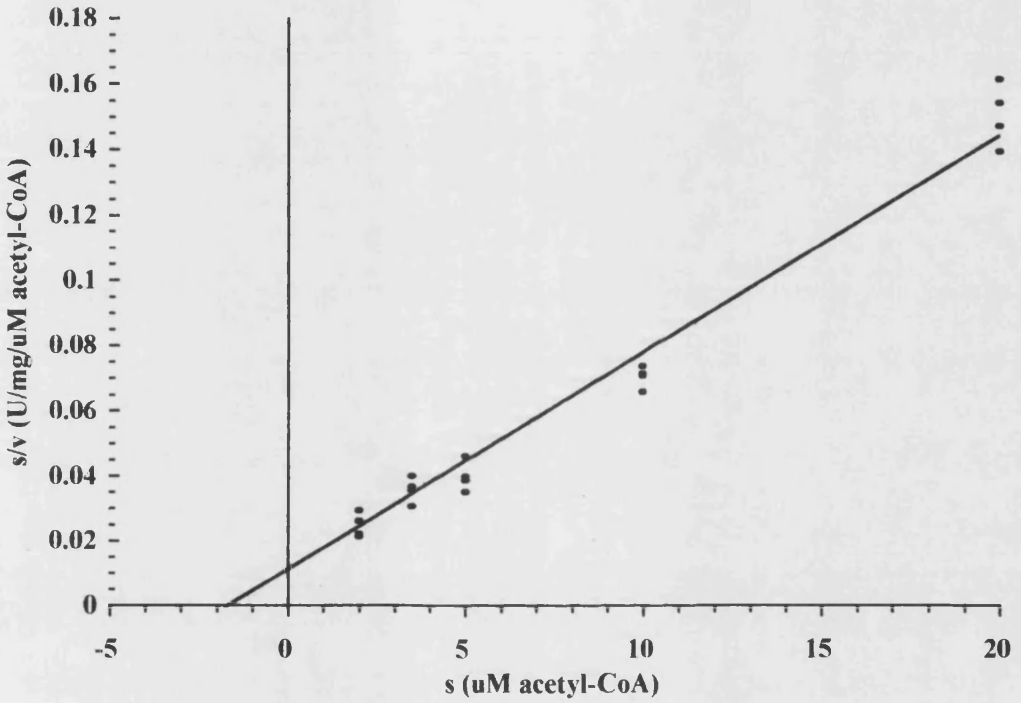


Figure 6.10 Hanes plot of the dependence of reaction rate on the concentration of acetyl-CoA for the recombinant *P. furiosus* citrate synthase. Each assay was performed at 55°C and contained 0.015 μg/ml enzyme and 0.2 mM oxaloacetate. The line is derived from the kinetic parameters obtained from a direct linear plot of the data (Table 6.2).

	Oxaloacetate		Acetyl-CoA	
	K_m^{app} (μM)	V_{max} (μmol/min/mg)	K_m^{app} (μM)	V_{max} (μmol/min/mg)
Direct linear plot	5.1	94.7	1.7	153.2
68% confidence limits	4.5 - 5.9	87.6 - 99.1	1.4 - 2.2	146.2 - 162.4

Table 6.2 Michaelis-Menten paramenters for the recombinant, *P. furiosus* citrate synthase, measured in 50 mM EPPS pH 8.0 (HCl), 0.2 mM EDTA, 100 mM KCl at 55°C.

6.3.6 Thermostability of the recombinant, *P. furiosus* citrate synthase

Under the conditions tested (30 μg pure enzyme per ml of 50 mM potassium phosphate pH 6.5 (20°C)), the recombinant citrate synthase had a half-life (50 % residual activity) of 20 min at 100°C (Figure 6.11). This decreased to 6 min and 1 min at 102°C and 104°C, respectively. An Arrhenius plot of ln k (first order rate constant for the loss of enzyme activity with time) as a function of temperature (1/T ($°K^{-1}$)) allows the comparison of the *P. furiosus* enzyme with the citrate synthases of other species. As shown in Figure 6.12, the *P. furiosus* enzyme is the most thermostable citrate synthase studied to date, followed by the enzyme from *Sulfolobus solfataricus* which was measured as a cell extract, rather than in a pure form. The effect of stabilising compounds that are found in hyperthermophilic organisms is discussed in Section 6.4.3

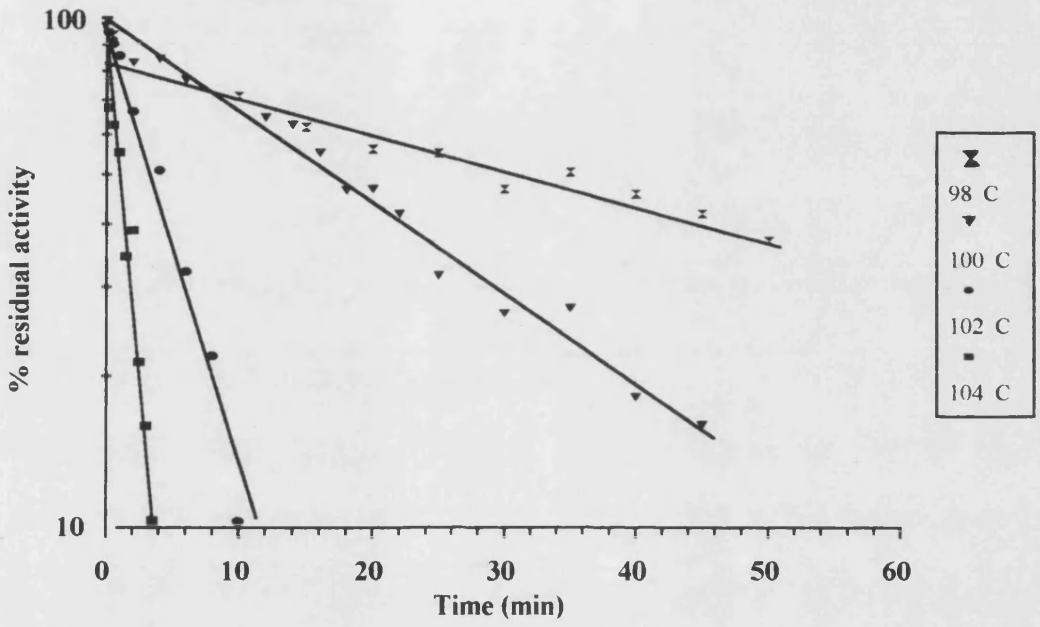


Figure 6.11 Thermal inactivation of the *P. furiosus* citrate synthase.

Volumes of 50 µl of pure enzyme at a concentration of 30 µg/ml were sealed in glass capillary tubes and incubated at the appropriate temperature. Each data point represents the percentage residual activity at each time point.

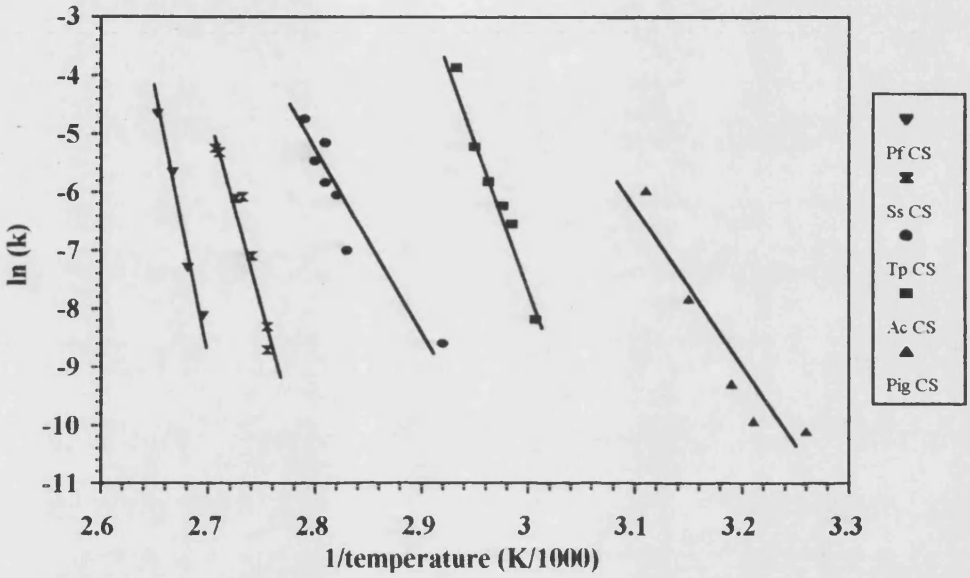


Figure 6.12 Arrhenius plot depicting the thermal inactivation of the citrate synthases of 5 species, pig [Russell, 1994], *A. anitratum* (Ac CS) [Shepherd, 1993], *T. acidophilum* (Tp CS) [Russell, 1994], *S. solfataricus* (Ss CS) [M. J. Danson, personal communication] and *P. furiosus* (Pf CS).

6.4 DISCUSSION

6.4.1 Expression of the *P. furiosus* citrate synthase gene in *E. coli* JM105

The citrate synthase gene of *P. furiosus* was expressed constitutively in *E. coli* JM105 despite being transcribed under the control of an IPTG-inducible *tac* promoter. This may be due to the promoter being 'leaky', thereby allowing a low rate of transcription to occur continuously, or the particular transformant selected for analysis, PfCS(EXP5), may not contain the *lacI*^{q} gene which encodes a repressor of the *tac* promoter. The strain of *E. coli* chosen for this expression study, *E. coli* JM105, is host for an F' episome which carries the *lacI*^{q} gene (Section 2.2.2). Loss of this episome would not affect the growth of the *E. coli* JM105 in complex media, but may explain the constitutive nature of gene expression. Alternatively, the amount of repressor produced from the *lacI*^{q} gene may not be adequate for complete repression of the *tac* promoter. Due to time constraints, the level of gene expression was not optimised. However, as low-level, constitutive expression produced a functional, hyperthermostable enzyme, this approach may be preferential to protocols involving induction.

6.4.2 Kinetic characterisation of the recombinant citrate synthase

The expression, in *E. coli* JM105, of the open reading frame determined in the preceding two chapters, confirmed that it encodes a functional, hyperthermostable enzyme. The enzyme was shown to be a homodimer with a subunit molecular weight of 42.6 ± 2.2 kDa; all archaeal citrate synthases studied to date have been shown to be dimeric.

The difficulties outlined in Chapter 3, involving low quantities of enzyme in *P. furiosus* cells, contaminated *P. furiosus* cell paste and possible proteolytic degradation of the enzyme did not allow an accurate determination of the Michaelis-Menten constants for the native enzyme. However, values of K_m^{app} and V_{max} for the recombinant enzyme were the same order of magnitude as determined for other citrate synthases although the K_m^{app} for both acetyl-CoA and oxaloacetate were of the lowest determined to date (Table 6.3). This may represent an increase in the affinity of the enzyme for the substrates which may be essential when considering the thermolability of oxaloacetate.

Organism	K_m^{app} oxaloacetate (μM)	K_m^{app} acetyl-CoA (μM)
Pig heart	10	7
Bacillus megaterium	12	76
Methanosarcina barkeri	18	7
Halobacterium halobium	58	211
Haloferax volcanii	38	75
Thermoplasma acidophilum	5	6
Sulfolobus acidocaldarius	20	10
Pyrococcus furiosus	5	2

Table 6.3 Comparison of the K_m^{app} values for oxaloacetate and acetyl-CoA for dimeric citrate synthases

6.4.3 Thermostability of the recombinant citrate synthase

A preliminary thermostability analysis demonstrated that the half-life of the recombinant, *P. furiosus* citrate synthase under the conditions tested was 20 min at 100°C. This is comparable to the half-life determined for GAPDH of *P. woesei* ($t_{½}$ = 44 min at 100°C), measured under anaerobic conditions [Zwickl *et al.*, 1990]. Although the *P. furiosus* enzyme represents the most thermostable citrate synthase studied to date, from the stability data presented here (Section 6.3.6), the ability of the enzyme to function at 104°C (the upper growth temperature of *P. furiosus*) is questionable. However, it has been proposed by Hensel and König [1988] that it may be of advantage to a thermophilic organism to be able to increase the thermostability of its intracellular enzymes by adjusting extrinsic conditions, such as ionic strength. This would allow the organism to compete over a greater temperature range than if the enzyme had an intrinsic, optimum thermostability at an elevated temperature yet was catalytically inefficient at lower temperatures. Thus, a combination of intrinsic thermostability with extrinsic support would allow maximum flexibility of growth conditions.

The intracellular environment of *Pyrococcus woesei* contains 600 mM K^+, countered by the novel anion, di-myo-inositol-1,1'-phosphate (DIP) [Scholz *et al.*, 1992]; this has been shown to increase the thermostability of the *P. woesei* GAPDH, *in vitro*. Another unusual anion has been isolated from the hyperthermophilic methanogen, *Methanothermus fervidus* (T_{opt} = 83°C), namely, cyclic 2,3-diphospho-glycerate (cDPG) [Hensel and König, 1988; Hensel and Jakob, 1994]. This also has a thermostabilising effect on enzymes *in vitro* and, moreover, the intracellular concentration of this trivalent anion has been shown to increase with the growth temperature of the organism up to the optimal growth temperature. The decrease in

cDPG concentration above the optimum growth temperature may be due to the decrease in the number of viable cells. *In vitro*, potassium phosphate has been shown to have a similar stabilising effect as cDPG. However, use of an organic phosphate compound may be advantageous to the organism in that its synthesis can be closely controlled.

A second compound, 2-O-α-D-mannopyranosylglycerate has been isolated from *P. woesei* and shown to stabilise *P. woesei* GAPDH *in vitro* [Hensel *et al.*, 1993] This compound may have an analogous role to the thermostabilising effects of polyols such as erythritol, xylitol and sorbitol [Lozano *et al.*, 1994 and references therein]. These compounds, which have a high hydrogen bonding capacity, are thought to increase protein thermostability by lowering the concentration of 'free' water and thereby increasing the probability that an intramolecular hydrogen bond in the protein remains intact. This effect can be described as lowering the water activity (A_w), which is defined as the ratio of water vapour pressure over a medium to that over pure water. Using bovine α-chymotrypsin as a model enzyme, Lozano *et al.* [1994] showed that the addition of sorbitol to a concentration of 4M increased the thermostability of the enzyme at 60°C by a factor of 700. More impressively, a thermostable xylanase, isolated from a species of *Thermotoga* (T_{opt} = 85°C), was incubated in 100 % (v/v) sorbitol at 120°C for 10 min and retained 85 % of its activity [Daniel R., personal communication]. Thus, extrinsic factors can have a significant effect on protein thermostability.

CHAPTER 7

ANALYSIS OF THE GENE SEQUENCE AND DERIVED PROTEIN SEQUENCE OF THE *P. FURIOSUS* CITRATE SYNTHASE

7.0 INTRODUCTION

The preceding chapter has confirmed that the gene sequence elucidated in Chapters 4 and 5 encodes a functional, hyperthermostable citrate synthase. It is now essential to analyse the sequence in order to fulfil the project aims outlined in Section 1.5. Therefore, this chapter describes the methods used and inferences obtained, regarding the gene sequence, the phylogenetic position of *P. furiosus* and the molecular basis of hyperthermostability.

7.1 METHODS

7.1.1 Codon usage

The nucleotide sequence generated in Chapters 4 and 5 was translated into amino acid sequence using STADEN PLUS Version 6 (Amersham International plc.). This program also generated a codon usage table from the citrate synthase open reading frame.

7.1.2 Sequence analysis using GCG

The primary sequence was analysed by the following programs of the GCG Sequence Analysis Software package [Devereux *et al.*, 1984]. The amino acid composition of known citrate synthase sequences was calculated using the COMPOSITION program. The sequences of the *C. elegans* and *N. crassa* citrate synthase were omitted from this analysis as their primary sequences are derived from the DNA sequences and included a mitochondrial leader sequence.

The percentage sequence identity between each citrate synthase primary sequence was determined using the BESTFIT program, which constructs an optimum alignment of the best segment of similarity between two sequences. The primary sequences of citrate synthases were aligned using the PILEUP program. Default settings were used in all cases. To allow the incorporation of the *C. elegans* and the *N. crassa* sequences in the alignment, the mitochondrial leader sequences were deleted at an arbitrary position.

7.1.3 Phylogenetic analysis of the primary sequences of citrate synthases

The multiple sequence alignment generated by the PILEUP program [University of Wisconsin GCG] (Section 7.1.2) was analysed further using PHYLIP (Phylogeny Inference Package) Version 3.5c [Felsenstein, 1993]. The concepts and methods of phylogenetic inference are introduced in Section 1.2. A distance measure between each pair of sequences of the multiple sequence alignment was computed using the PROTDIST program, which scores each amino acid substitution according to the type of residue exchange (conservative or non-conservative) and the underlying nucleotide changes required to effect the replacement. A FITCH analysis was performed on the distance matrix of citrate synthase sequences which computes the most probable phylogenetic tree without assuming a molecular clock.

To assess the variability in the sequence data, a resampling procedure, the bootstrap analysis [reviewed in Felsenstein, 1988], was applied using the SEQBOOT program. Each resampled data set was then analysed by PROTDIST, followed by FITCH; 100 replicates were analysed in this manner.

7.1.4 Structural alignment

A structural alignment of the primary sequence of the *P. furiosus* citrate synthase with those of pig and *T. acidophilum* was generated by R. J. M. Russell (University of Bath) using the program AMPS [Barton and Sternberg, 1990]. The pig and *T. acidophilum* sequences were aligned in a manner directly related to their secondary structure, as determined by X-ray crystallography. The *P. furiosus* sequence was added to this alignment by AMPS, which constrains the alignment by increasing the gap penalty within regions of secondary structure.

7.1.5 Homology modelling of the *P. furiosus* citrate synthase

In collaboration with R. J. M. Russell (University of Bath), residues 9 to 364 of the primary sequence of the *P. furiosus* citrate synthase were homology-modelled to the crystal structure of the *T. acidophilum* enzyme at 2.5Å resolution using the program O [Jones *et al.*, 1991]. It was not possible to model the amino- and carboxyl-termini of the *P. furiosus* sequence as the electron density of the *T. acidophilum* citrate synthase crystal structure was poorly defined in these areas. The sequence of the *T. acidophilum* monomer was changed for the *P. furiosus* sequence using the *mutate_replace* command to effect the desired substitutions. After each substitution, the *Lego_sidechain* command was used to display the energetically most favourable orientations that the new side chain could adopt (rotamers). In most cases, the rotamer that appeared the most frequently in known structures was selected, unless an inappropriate contact was

observed. Small insertions and deletions were built into the model using the commands *mutate_insert* and *mutate_delete*. Possible loop conformations at the site of the insertion or deletion were displayed using *Lego_loop*, *Lego_mainchain* and *Lego_ca* commands. As with the selection of rotamers, the most appropriate loop conformation was selected by eye, from a database of loop conformations from known crystal structures, although the residues comprising the displayed loop were taken into consideration.

The dimeric form of the modelled enzyme was created by retrieving the co-ordinates of the second subunit of the *T. acidophilum* citrate synthase and using the *Lsq* subset of commands to perform a least squares superimposition of the model onto the second subunit. The modelled dimer was then energy minimised using XPLOR (Molecular Simulations Inc.) [Brünger, 1990]. Energy minimisation is a cyclical process whereby a composite energy value, based on factors such as bond lengths, dihedral angles and electrostatic, van der Waals and hydrogen bond interactions, is reduced as the conformation of the overall molecule is adjusted to an eventual conformation of minimal energy. For the homology-modelled citrate synthase, 200 cycles of energy minimisation were applied.

A number of programs are available to assess the stereochemical quality of the homology-modelled structure. Initially, the program DSSP [Kabsch and Sander, 1983] was used to define the secondary structure and solvent exposure of the model from the atomic co-ordinates generated after the energy minimisation procedure. The DSSP output was then analysed by the program ENVIRONMENTS followed by VERIFY which assess the acceptibility of the environment of each residue. In addition, the co-ordinates of the model were applied to the program PROCHECK [Morris *et al.*, 1992], which computed the main chain parameters and side chain angles for the modelled protein. A Ramachandran plot of the PROCHECK output demonstrated which residues had unfavourable stereochemistry; these were then altered using the program O and the energy minimisation and subsequent analysis repeated.

Citrate synthase is a predominantly α-helical protein. Therefore, as regular access to the model was not possible, the DSSP output from the eventual structure was used to generate helical wheel diagrams using the program HERA [Hutchinson and Thornton, 1990]. In combination with both the structural alignment and the schematic representation of the *T. acidophilum* citrate synthase monomer (Figure 7.5), these were used to illustrate the *P. furiosus* primary sequence.

7.2 RESULTS AND DISCUSSION

7.2.1 The gene sequence of the *P. furiosus* citrate synthase gene

The PCR-mediated approach, described in Chapters 4 and 5, resulted in the determination of the nucleotide sequence of a 1.3 kb segment of *P. furiosus* genomic DNA (Figure 7.1). Of this, 1134 bp comprised the citrate synthase coding region, commencing with a Met start codon and ending with a translation termination codon (TAG). Upstream and downstream of the coding region, 85 bp of nucleotide sequence were elucidated in search of regulatory sequences for the transcription and translation processes.

From an alignment of the transcription start sites of a number of archaeal genes, an archaeal promoter sequence has been formulated [Zillig *et al.*, 1988]. This comprises two sequences, box B (A/T T G A/C), which is located at the transcription start site, and box A (T T T A A/T T), which is located approximately 26 bp upstream of box B. Box A is more conserved than box B. On assessing the region upstream of the citrate synthase coding region, neither of these consensus sequences was obvious, although an A-T rich region was noted at -47 to -54 (inclusive) and the 4 nucleotide

sequence ACGA at -17 to -20 (inclusive). Initiation of transcription at this site would generate a transcript that included a putative ribosome binding sequence, GAGGT, at -2 to -6 (inclusive). This sequence shows homology to the 3' terminus of the 16S rRNA of the related Archaeon, *Thermococcus celer* [Achenbach-Richter, 1988].

Thus, the upstream region of the *P. furiosus* citrate synthase gene shows divergence from the consensus archaeal signal sequences. Alternatively, the lack of defined signals may reflect a polycistronic mode of gene organisation whereby a single mRNA transcript is generated, from which translation is initiated at multiple sites. However, no open reading frames were observed in the regions immediately upstream or downstream of the *P. furiosus* citrate synthase gene. The sequence of a 3.1 kb portion of *P. woesei* genomic DNA was elucidated by Zwickl *et al.* [1990] in search of the gene encoding GAPDH. A total of five open reading frames were identified in the regions adjacent to the GAPDH gene, but no homology to other catabolic enzymes was observed. Hence, co-ordinated gene expression was not inferred in this organism.

The signal for transcription termination has been suggested to comprise stretches of 4-5 pyrimidine nucleosides [Zillig *et al.*, 1988]. Of the 85 bp elucidated downstream of the citrate synthase coding region, a stretch of 5 pyrimidine nucleosides was observed at 1152 to 1156 (inclusive) that was preceded by a stem and loop structure at 1139 to 1148 (inclusive). These may serve to terminate transcription.

The genome composition of *P. furiosus* was determined by Fiala and Stetter [1986] to be 38 % G+C. The nucleotide sequence generated here agreed with this estimate (39 % G+C) and strong bias against guanosine and cytosine was observed in the codon usage table derived from the citrate synthase coding region (Table 7.1). For example, the GGG codon for Gly was not used in this coding region, neither was the CCG codon for Pro nor the codons commencing with CG for Arg. The third phenomenon was also observed in the *P. woesei* GAPDH coding region.

Figure 7.1 Nucleotide sequence of the *P. furiosus* citrate synthase gene

Non-coding sequence is shown in lower case; coding sequence is shown in upper case. The ATG start codon and translation termination codon are shown in bold font.

```
catatgtccacaaatgttgaacattgtcccgatatttttgcaatatcattataaataatatctctacgaaacatt
   -80       -70       -60       -50       -40       -30       -20
gtatacaggtgtttacaacttgtaacagggctataaaaacgttatagtaatatttattatagagatgctttgtaa

                1                                  10                                  20
                N  T  E  K  Y  L  A  K  G  L  E  D  V  Y  I  D  Q  T  N  I  C
ttcagaggtg**ATG**AATACGGAAAAATACCTTGCTAAAGGTCTAGAAGATGTCTACATAGATCAAACAAATATATG
          1       10        20        30        40        50        60
aagtctccacTACTTATGCCTTTTTATGGAACGATTTCCAGATCTTCTACAGATGTATCTAGTTTGTTTATATAC

                          30                                  40
  Y  I  D  G  K  E  G  K  L  Y  Y  R  G  Y  S  V  E  E  L  A  E  L  S  T  F
CTATATCGATGGAAAAGAGGGAAAGCTATACTACAGAGGATACAGTGTAGAAGAGCTAGCAGAGTTAAGCACGTT
   70        80        90        100       110       120       130       140
GATATAGCTACCTTTTCTCCCTTTCGATATGATGTCTCCTATGTCACATCTTCTCGATCGTCTCAATTCGTGCAA

          50                                  60                                  70
  E  E  V  V  Y  L  L  W  W  G  K  L  P  S  L  S  E  L  E  N  F  K  K  E  L
TGAGGAAGTGGTGTATCTCCTCTGGTGGGGAAAATTACCATCCCTAAGTGAACTAGAGAATTTCAAAAAAGAGCT
       150       160       170       180       190       200       210
ACTCCTTCACCACATAGAGGAGACCACCCCTTTTAATGGTAGGGATTCACTTGATCTCTTAAAGTTTTTTCTCGA

                          80                                  90
  A  K  S  R  G  L  P  K  E  V  I  E  I  M  E  A  L  P  K  N  T  H  P  M  G
AGCAAAAAGTAGAGGCCTTCCAAAAGAAGTTATAGAAATAATGGAAGCACTACCTAAAAATACTCACCCAATGGG
  220       230       240       250       260       270       280       290
TCGTTTTTCATCTCCGGAAGGTTTTCTTCAATATCTTTATTACCTTCGTGATGGATTTTTTAGAGTGGGTTACCC

         100                                 110                                 120
  A  L  R  T  I  I  S  Y  L  G  N  I  D  D  S  G  D  I  P  V  T  P  E  E  V
CGCTTTAAGAACCATAATCTCATATCTAGGAAACATTGACGATAGTGGAGACATTCCAGTAACCCCTGAAGAAGT
       300       310       320       330       340       350       360
GCGAAATTCTTGGTATCAGAGTATAGATCCTTTGTAACTGCTATCACCTCTGTAAGGTCATTGGGGACTTCTTCA

                          130                                 140
  Y  R  I  G  I  S  V  T  A  K  I  P  T  I  V  A  N  W  Y  R  I  K  N  G  L
TTACAGGATAGGAATTAGTGTAACTGCAAAGATACCAACAATAGTAGCTAATTGGTATAGAATTAAGAATGGTCT
  370       380       390       400       410       420       430       440
AATGTCCTATCCTTAATCACATTGACGTTTCTATGGTTGTTATCATCGATTAACCATATCTTAATTCTTACCAGA

         150                                 160                                 170
  E  Y  V  P  P  K  E  K  L  S  H  A  A  N  F  L  Y  M  L  H  G  E  E  P  P
CGAATATGTTCCTCCAAAAGAAAAACTTAGTCATGCAGCAAATTTCCTATACATGCTCCACGGTGAAGAACCCCC
       450       460       470       480       490       500       510
GCTTATACAAGGAGGTTTTCTTTTTGAATCAGTACGTCGTTTAAAGGATATGTACGAGGTGCCACTTCTTGGGGG
```

```
                        180                             190
  K  E  W  E  K  A  M  D  V  A  L  I  L  Y  A  E  H  E  I  N  A  S  T  L  A
CAAAGAATGGGAGAAAGCCATGGATGTGGCCCTAATTCTATATGCTGAACATGAAATTAACGCATCGACTTTAGC
  520       530       540       550       560       570       580       590
GTTTCTTACCCTCTTTCGGTACCTACACCGGGATTAAGATATACGACTTGTACTTTAATTGCGTAGCTGAAATCG

     200                             210                             220
  V  M  T  V  G  S  T  L  S  D  Y  Y  S  A  I  L  A  G  I  G  A  L  K  G  P
AGTAATGACCGTTGGCTCTACTCTTAGCGACTACTATTCAGCCATCCTCGCGGGAATAGGAGCTTTAAAGGGCCC
       600       610       620       630       640       650       660
TCATTACTGGCAACCGAGATGAGAATCGCTGATGATAAGTCGGTAGGAGCGCCCTTATCCTCGAAATTTCCCGGG

                        230                             240
  I  H  G  G  A  V  E  E  A  I  K  Q  F  M  E  I  G  S  P  E  K  V  E  E  W
AATCCATGGAGGTGCAGTAGAGGAGGCCATAAAACAGTTTATGGAAATAGGATCTCCCGAGAAAGTAGAAGAATG
  670       680       690       700       710       720       730       740
TTAGGTACCTCCACGTCATCTCCTCCGGTATTTTGTCAAATACCTTTATCCTAGAGGGCTCTTTCATCTTCTTAC

     250                             260                             270
  F  F  K  A  L  Q  Q  K  R  K  I  M  G  A  G  H  R  V  Y  K  T  Y  D  P  R
GTTCTTTAAGGCTCTTCAGCAGAAGAGAAAAATAATGGGTGCTGGTCATAGGGTGTATAAAACCTATGATCCAAG

       750       760       770       780       790       800       810
CAAGAAATTCCGAGAAGTCGTCTTCTCTTTTTATTACCCACGACCAGTATCCCACATATTTTGGATACTAGGTTC

                        280                             290
  A  R  I  F  K  K  Y  A  S  K  L  G  D  K  K  L  F  E  I  A  E  R  L  E  R
GGCTAGAATATTTAAAAAATATGCCTCCAAATTAGGAGATAAAAAGCTCTTCGAGATAGCTGAAAGATTAGAGAG
  820       830       840       850       860       870       880       890
CCGATCTTATAAATTTTTTATACGGAGGTTTAATCCTCTATTTTTCGAGAAGCTCTATCGACTTTCTAATCTCTC

     300                             310                             320
  L  V  E  E  Y  L  S  K  K  G  I  S  I  N  V  D  Y  W  S  G  L  V  F  Y  G
GTTAGTAGAAGAATACCTAAGCAAAAAAGGAATTAGTATAAATGTTGACTACTGGTCAGGATTAGTTTTCTATGG
       900       910       920       930       940       950       960
CAATCATCTTCTTATGGATTCGTTTTTTCCTTAATCATATTTACAACTGATGACCAGTCCTAATCAAAAGATACC

                        330                             340
  M  K  I  P  I  E  L  Y  T  T  I  F  A  M  G  R  I  A  G  W  T  A  H  L  A
AATGAAGATTCCAATAGAGCTTTACACAACAATATTTGCAATGGGTAGAATTGCGGGATGGACTGCTCATTTAGC
  970       980       990      1000      1010      1020      1030      1040
TTACTTCTAAGGTTATCTCGAAATGTGTTGTTATAAACGTTACCCATCTTAACGCCCTACCTGACGAGTAAATCG

     350                             360                             370
  E  Y  V  S  H  N  R  I  I  R  P  R  L  Q  Y  V  G  E  I  G  K  K  Y  L  P
TGAGTATGTTTCTCACAATAGAATTATTAGACCCAGGTTGCAGTATGTAGGAGAAATAGGAAAGAAGTACCTACC
      1050      1060      1070      1080      1090      1100      1110
ACTCATACAAAGAGTGTTATCTTAATAATCTGGGTCCAACGTCATACATCCTCTTTATCCTTTCTTCATGGATGG

           376
  I  E  L  R  R  *
CATAGAATTAAGGAGGTAGataaaatgagcattagacttccaccaagaaggaaaagttattgaagttgtcaaccg
 1120      1130      1140      1150      1160      1170      1180      1190
GTATCTTAATTCTTCCATCtattttactcgtaatctgaaggtggttcttccttttcaataacttcaacagttggc

gctctctaaagaccctaatgagccca
      1200      1210
cgagagatttctgggattactcgggt
```

Amino acid	Codon	Fraction	Amino acid	Codon	Fraction
Ala	GCG	0.07	Leu	TTG	0.03
Ala	GCA	0.36	Leu	TTA	0.30
Ala	GCT	0.39	Leu	CTG	0.00
Ala	GCC	0.18	Leu	CTA	0.35
			Leu	CTT	0.16
Arg	AGG	0.41	Leu	CTC	0.16
Arg	AGA	0.59			
Arg	CGG	0.00	Lys	AAG	0.33
Arg	CGA	0.00	Lys	AAA	0.67
Arg	CGT	0.00			
Arg	CGC	0.00	Met	ATG	1.00
Asn	AAT	0.82	Pro	CCG	0.00
Asn	AAC	0.18	Pro	CCA	0.53
			Pro	CCT	0.18
Asp	GAT	0.64	Pro	CCC	0.29
Asp	GAC	0.36			
			Phe	TTT	0.50
Cys	TGT	0.00	Phe	TTC	0.50
Cys	TGC	1.00			
			Ser	AGT	0.37
Gln	CAG	0.80	Ser	AGC	0.16
Gln	CAA	0.20	Ser	TCG	0.05
			Ser	TCA	0.16
Glu	GAG	0.36	Ser	TCT	0.16
Glu	GAA	0.64	Ser	TCC	0.10
Gly	GGG	0.00	Thr	ACG	0.13
Gly	GGA	0.62	Thr	ACA	0.27
Gly	GGT	0.24	Thr	ACT	0.33
Gly	GGC	0.14	Thr	ACC	0.27
His	CAT	0.38	Trp	TGG	1.00
His	CAC	0.62			
			Tyr	TAT	0.52
Ile	ATA	0.58	Tyr	TAC	0.48
Ile	ATT	0.33			
Ile	ATC	0.09	Val	GTG	0.18
			Val	GTA	0.41
End	TGA	0.00	Val	GTT	0.32
	TAA	0.00	Val	GTC	0.09
	TAG	1.00			

Table 7.1 Codon usage table constructed from the coding region of the *P. furiosus* citrate synthase gene

7.2.2 Alignment of the primary sequences of known citrate synthases

An alignment of the primary sequences of known citrate synthases is illustrated in Figure 7.2. The conservation of the active site His residues at positions 309 and 357 of the multiple sequence alignment demonstrates that the catalytic mechanism is likely to be conserved across the eukaryal, bacterial and archaeal domains. The third residue implicated in the catalytic mechanism is the Asp at position 417 of the multiple sequence alignment. This residue is conserved in all citrate synthases except the citrate synthase (CitA) of *B. subtilis* which has a functionally-conserved Glu at this position. These residues are equivalent to His 274, His 320 and Asp 375 of the pig heart enzyme (Section 1.1.3).

The multiple sequence alignment also highlights the recessed amino-terminus of the enzyme from Archaea and Gram-positive Bacteria. The amino-terminal extension of the pig heart enzyme has been shown in the crystal structure to form an α-helix (helix A) and a long loop region that rests in a groove on the surface of the second subunit. In addition to the recessed amino-terminus, a number of conserved gaps in the alignment occur in the sequences of the citrate synthases of Archaea and Bacteria.

7.2.3 Protein-derived phylogenetic analysis of the citrate synthase sequences

The multiple sequence alignment, shown in Figure 7.2, was used to generate a distance matrix from which a phylogenetic tree was calculated. Figure 7.3 illustrates the unrooted tree generated by this analysis, which indicates that the citrate synthase sequences cluster into three groups representing the sequences of the Eukarya, Gram-negative Bacteria and Gram-positive Bacteria/Archaea.

Figure 7.2 Alignment of the primary sequences of known citrate synthases created using PILEUP (University of Wisconsin GCG Sequence Analysis Software). Each sequence is referenced in Table 1.1. Residues that are conserved throughout all citrate synthase sequences are marked with an asterix

```
                        1                                                 50                *                 *                100
                Chicken ..........  .......ASS TNLKDVLAAL IPKEQARIKT FRQQHGGTAL GQITV.DMSY GGMRGMKGLV YETSVLDPDE G.IRFRGFSI PECQKLLPKG
                    Pig ..........  .......ASS TNLKDILADL IPKEQARIKT FRQQHGNTVV GQITV.DMMY GGMRGMKGLV YETSVLDPDE G.IRFRGYSI PECQKMLPKA
             C. elegans ..........  .....SAEGS TNLKEVLSKK IPAHNAKVKS FRTEHGSTVV QNVNI.DMIY GGMRSMKGMV TETSVLDPEE G.IRFRGYSI PECQKLLPKA
           Yeast(glyo) ..........  .......SQE KTLKERFSEI YPIHAQDVRQ FVKEHGKTKI SDVLL.EQVY GGMRGIPGSV WEGSVLDPED G.IRFRGRTI ADIQKDLPKA
           Yeast(mito) ..........  .......ASE QTLKERFAEI IPAKAQEIKK FKKEHGKTVI GEVLLEEQAY GGMRGIKGLV WEGSVLDPEE G.IRFRGRTI PEIQRELPKA
              N. crassa ..........  ......SSKT QTLKERFAEL LPENIEKIKA LRKEHGSKVV DKVTL.DQVY GGARGIKCLV WEGSVLDAEE G.IRFRGKTI PECQELLPKA
          Te. thermophila ..........  ........SQ TNLKKVIAEI IPQKQAELKE VKEKYGDKVV GQYTV.KQVI GGMRGMKGLM SDLSRCDPYQ G.IIFRGYTI PQLKEFLPKA
            A. anitratum ....SEATGK KAVLHLDGKE .IELPIYSGT LGPDVIDVKD VLAS.GHFTF DP........ .GFMATASCE SKITFIDGDK GILLHRGYPI DQ.L......
           Ps. aeruginosa .......ADK KAQLIIEGSA PVELPVLSGT MGPDVVDVRG LTAT.GHFTF DP........ .GFMSTASCE SKITYIDGDK GVLLHRGYPI EQ.L......
                E. coli .......ADT KAKLTLNGDT AVELDVLKGT LGQDVIDIRT LGSK.GVFTF DP........ .GFTSTASCE SKITFIDGDE GILLHRGFPI DQ.L......
              Ac. acetii SASQKEGKLS TATISVDGKS A.EMPVLSGT LGPDVIDIRK LPAQLGVFTF DP........ .GYGETAACN SKITFIDGDK GVLLHRGYPI AQ.L......
            Co. burnetii .......SNR KAKLSFENQS .VEFPIYSPT LGKDVIDVKT L.GNHGAYAL DV........ .GFYSTAACE SKITFIDGEK GILLYRGYPI DQ.L......
            R. prowazekii .TNGNNNNLE FAELKIRGK. LFKLPILKAS IGKDVIDISR VSAEADYFTY DP........ .GFMSTASCQ STITYIDGDK GILWYRGYDI KD.L......
        B. subtilis (CitZ) ..........  .......... .......... .......... ........TA TR........ .GLEGVVATT SSVSSI..ID DTLTYVGYDI DD.L......
              P. furiosus ..........  .......... .......... .......... ....NTEKYL AK........ .GLEDVYIDQ TNICYIDGKE GKLYYRGYSV EE.L......
        B. subtilis (CitA) ..........  .......... .......... .......... .........V HY........ .GLKGITCVE TSISHIDGEK GRLIYRGHHA KD.I......
B. coagulans (strain C4) ..........  .......... .......... .......... ....VNTNQF IP........ .GLEGVIASE TKISFLDTVN SEIVIKGYDL LA.L......
            M. smegmatis ..........  .......... .......... .........T TATESEAPRI HK........ .GLAGVVVDT TAISKVVPET NSLTYRGYPV QD.L......
            T. acidophilum ..........  .......... .......... .......... ....PETEEI SK........ .GLEDVNIKW TRLTTIDGNK GILRYGGYSV EDII......

                        101                                               150                                               200
                Chicken G......XGG EPLPEGLFWL LVTGQIPTGA QVSWLSKEWA KRAALPSHVV TMLDNFP.TN LHPMSQLSAA ITALNSESNF ARAYAEG.IL RTKYWEMVYE
                    Pig K......GGE EPLPEGLFWL LVTGQIPTEE QVSWLSKEWA KRAALPSHVV TMLDNFP.TN LHPMSQLSAA ITALNSESNF ARAYAEG.IH RTKYWELIYE
             C. elegans K......GGE EPLPEAIWWL LCTGDVPSEA QTAAITKEWN ARADLPTHVV RMLDNFP.DN LHPMAQFIAA IAALNNESKF AGAYARG.VA KASYWEYAYE
           Yeast(glyo) K......GSS QPLPEALFWL LLTGEVPTQA QVENLSADLM SRSELPSHVV QLLDNLP.KD LHPMAQFSIA VTALESESKF AKAYAQG.IS KQDYWSYTFE
           Yeast(mito) E......GST EPLPEALFWL LLTGEIPTDA QVKALSADLA ARSEIPEHVI QLLDSLP.KD LHPMAQFSIA VTALESESKF AKAYAQG.VS KKEYWSYTFE
              N. crassa P......GGK EPLPEGLFWL LLTGEVPSEQ QVRDLSAEWA ARSDVPKFIE ELIDRCP.SD LHPMAQLSLA VTALEHTSSF ARAYAKG.IN KKEYWGYTFE
          Te. thermophila DPKAADQANQ EPLPEGIFWL LMTGQLPTHA QVDALKHEWQ NRGTVNQDCV NFILNLP.KD LHSMTMLSMA LLYLQKDSKF AKLYDEGKIS KKDYWEPFYE
            A. anitratum ........AT QADYLETCYL LLNGELPTAE QKVEFDAKVR AHTMVHDQVS RFFNGFR.RD AHPMAIMVGV VGALSAFYHN NLDIEDI... .....NHREI
           Ps. aeruginosa ........AE KSDYLETCYL LLNGELPTAA QKEQFVGTIK NHTMVHEQLK TFFNGFR.RD AHPMAVMCGV IGALSAFYHD SLDITNP... .....KHRQV
                E. coli ........AT DSNYLEVCYI LLNGEKPTQE QYDEFKTTVT RHTMIHEQIT RLFHAFR.RD SHPMAVMCGI TGALAAFYHD SLDVNNP... .....RHREI
              Ac. acetii ........DE NASYEEVIYL LLNGELPNKV QYDTFTNTLT NHTLLHEQIR NFFNGFR.RD AHPMAILCGT VGALSAFYPD ANDIAIP... .....ANRDL
            Co. burnetii ........AD KSDYMEVCYL LMYGELPNKG EKEKFVRTIK EHTSVYEQVT KFFNGFH.YD AHPMAMVLST IGALSAFYHD ALDITKP... .....ADREL
            R. prowazekii ........AE KSDFLEVAYL MIYGELPSSD QYCNFTKKVA HHSLVNERLH YLFQTFC.SS SHPMAIMLAA VGSLSAFYPD LLNFNET... .....DY.EL
        B. subtilis (CitZ) ........TE NRSFEEIIYL LWHLRLPNKK ELEELKQQLA KEAAVPQEII EHFKSYSLEN VHPMAALRTA ISLLGLLDSE AD.TMNP... .....EANYR
              P. furiosus ........AE LSTFEEVVYL LWWGKLPSLS ELENFKKELA KSRGLPKEVI EIMEALP.KN THPMGALRTI ISYLGNIDDS GDIPVTP... .....EEVYR
        B. subtilis (CitA) ........AL NHSFEEAAYL ILFGKLPSTE ELQVFKDKLA AERNLPEHIE RLIQSLP.NN MDDMSVVRTV VSALG..... .ENTYTF... .....HPKTE
B. coagulans (strain C4) ........SK TKGYLDIVHL LLEGTIPNEA EKQHLEETLK QEYDVPDEII QVLSLLP.KT AHPMDALRTG VSVLASFDTE L.LNREH... .....STNLK
            M. smegmatis ........AA QCSFEQ.VYL LWHGELP.TD QLALFSQRER ASRRIDRSMQ ALLAKLP.DN CHPMDVVRTA ISYLGAEDLE EDVD.TA... .....EANYA
            T. acidophilum ........AS GAQDEEIQYL FLYGNLPTEQ ELRKYKETVQ KGYKIPDFVI NAIRQLP.RE SDAVAMQMAA VAAMAA..SE TKFKWNK... .....DTDRD
```

```
                    201                                                 250                                                 300
Chicken             SAMDLIAKLP CVAAKIYRNL YRAGSSIGAI DSKLDWSHNF TNML..... .GY.TDAQFT ELMRLYLTIH SDHEGGNVSA HTSHLVGSAL SDPYLSFAAA
Pig                 DCMDLIAKLP CVAAKIYRNL YREGSSIGAI DSKLDWSHNF TNML..... .GY.TDAQFT ELMRLYLTIH SDHEGGNVSA HTSHLVGSAL SDPYLSFAAA
C. elegans          DSMDLLAKLP TVAAIIYRNL YRDGSAVSVI DPKKDWSANF SSML..... .GY.DDPLFA ELMRLYLVIH SDHEGGNVSA HTSHLVGSAL SDPYLSFSAA
Yeast(glyo)         DSLDLLGKLP VIAAKIYRNV FKDGK.MGEV DPNADYAKNL VNLI..... .GS.KDEDFV DLMRLYLTIH SDHEGGNVSA HTSHLVGSAL SSPYLSLASG
Yeast(mito)         DSLDLLGKLP VIASKIYRNV FKDGK.ITST DPNADYGKNL AQLL..... .GY.ENKDFI DLMRLYLTIH SDHEGGNVSA HTTHLVGSAL SSPYLSLAAG
N. crassa           DSMDLIAKLP TIAARIYQNV FKGGK.VAAV QKDKDYSFNF ANQL..... .GFGDNKDFV ELLRLYLTIH TDHEGGNVSA HTTHLVGSAL SSPFLSVAAG
Te. thermophila     DSMDLIAKIP RVAAIIYRHK YRDSKLIDS. DSKLDWAGNY AHMM..... .GF.EQHVVK ECIRGYLSIH CDHEGGNVSA HTTHLVGSAL SDPYLSYSAG
A. anitratum        TAIRLIAKIP TLAAWSYK.. YTVGQPFIYP RNDLNYAENF LHMMFATPAD RDYKVNPVLA RAMDRIFTLH ADHE.QNAST STVRLAGSTG ANPYACISAG
Ps. aeruginosa      SAHRLIAKMP TIAAMVYK.. YSKGEPMMYP RNDLNYAENF LHMMFNTPCE TK.PISPVLA KAMDRIFILH ADHE.QNAST STVRLAGSSG ANPFACIASG
E. coli             AAFRLLSKMP IMAAMCYK.. YSIGQPFVYP RNDLSYAGNF LNMMFSTPCE P.YEVNPILE RAMDRILILH ADHE.QNAST STVRTAGSSG ANPFACIAAG
Ac. acetii          AAMRLIAKIP TIAAWAYK.. YTQGEAFIYP RNDLNYAENF LSMMFARMSE P.YKVNPVLA RAMNRILILH ADHE.QNAST STVRLAGSTG ANPFACIAAG
Co. burnetii        SAIRLIAKMP TLAAMSYK.. YSIGQPFMHP RRAMNYAENF LHMLFGTPYE ET.EPDPVLA RAMDRIFILH ADHE.QNAST TTVRVAGSTG ANPFACISAG
R. prowazekii       TAIRMIAKIP TIAAMSYK.. YSIGQPFIYP DNSLDFTENF LHMMFATPC. TKYKVNPIIK NALNKIFILH ADHE.QNAST STVRIAGSSG ANPFACISTG
B. subtilis (CitZ)  KAIRLQAKVP GLVAAFSR.. IRKGLEPVEP REDYGIAENF LYTLNGE... ...EPSPIEV EAFNKALILH ADHE.LNAST FTARVCVATL SDIYSGITAA
P. furiosus         IGISVTAKIP TIVANWYR.. IKNGLEYVPP KEKLSHAANF LYMLHGE... ...EPPKEWE KAMDVALILY AEHE.INAST LAVMTVGSTL SDYYSAILAG
B. subtilis (CitA)  EAIRLIAITP SIIAYRKR.. WTRGEQAIAP SSQYGHVENY YYMLTGE... ...QPSEAKK KALETYMILA TEHG.MNAST FSARVTLSTE SDLVSAVTAA
B. coagulans (strain C4) RAYQLLGKIP NIVANSYH.. ILHSEEPVQP LQDLSYSANF LYMITGK... ...KPTELEE KIFDRSLVLY SEHE.LPNST FTARVIASTL SDLYGALTGA
M. smegmatis        KSLRMFAVLP TIVATDIR.. RRQGLTPIPP HSQLGYAQNF LNMCFGE... ...VPEPVVV RAFEQSMVLY AEHS.FNAST FAARVVTSTQ SDIYSAVTAA
T. acidophilum      VAAEMIGRMS AITVNVYR.. HIMNMPAELP KPSDSYAESF LNAAFGR... ...KATKEEI DAMNTALILY TDHE.VPAST TAGLVAVSTL SDMYSGITAA

                    301  *  **  *                                         350    *  * *      * *                                 400
Chicken             MNGLAGPLHG LANQEVLGWL AQLQKAXXXA GADASLRDYI WNT..LNSGR VVPGYGHAVL RKT.DPRYTC QREFALKHLP G....DPMFK LVAQLYKIVP
Pig                 MNGLAGPLHG LANQEVLVWL TQLQKEVGKD VSDEKLRDYI WNT..LNSGR VVPGYGHAVL RKT.DPRYTC QREFALKHLP H....DPMFK LVAQLYKIVP
C. elegans          MAGLAGPLHG LANQEVLVFL NKIVGEIGFN YTEEQLKEWV WKH..LKSGQ VVPGYGHAVL RKT.DPRYEC QREFALKHLP N....DDLFK LVSTLYKITP
Yeast(glyo)         LNGLAGPLHG RANQEVLEWL FALKEEVNDD YSKDTIEKYL WDT..LNSGR VIPGYGHAVL RKT.DPRYMA QRKFAMDHFP D....YELFK LVSSIYEVAP
Yeast(mito)         LNGLAGPLHG RANQEVLEWL FKLREEVKGD YSKETIEKYL WDT..LNAGR VVPGYGHAVL RKT.DPRYTA QREFALKHFP D....YELFK LVSTIYEVAP
N. crassa           LNGLAGPLHG LANQEVLNWL TEMKKVIGDD LSDEAITKYL WDT..LNAGR VVPGYAHAVL RKT.DPRYSA QRKFAQEHLP E....DPMFQ LVSQVYKIAP
Te. thermophila     VNGLAGPLHG LANQEVLKWL LQFIEEKGTK VSDKDIEDYV DHV..ISSGR VVPGYGHAVL RDT.DPRFHH QVDFSKFHLK D....DQMIK LLHQCADVIP
A. anitratum        ISALWGPAHG GANEAVLKML DEIG..... .SVENVAEFM EKVKRKE..V KLMGFGHRVY .KNFDPRAKV MKQTCDEVLE ALG.IN.DPQ LALAMELERI
Ps. aeruginosa      IAALWGPAHG GANEAVLRML DEIG..... .DVSNIDKFV EKAKDKNDPF KLMGFGHRVY .KNFDPRAKV MKQTCDEVLQ ELG.IN.DPQ LELAMKLEEI
E. coli             IASLWGPAHG GANEAALKML EEIS..... .SVKHIPEFF RRAKDKNDSF RLMGFGHRVY .KNYDPRATV MRETCHEVLK ELG.TK.DDL LEVAMELENI
Ac. acetii          IAALWGPAHG GANEAVLKML ARIG..... .KKENIPAFI AQVKDKNSGV KLMGFGHRVY .KNFDPRAKI MQQTCHEVLT ELG.IKDDPL LDLAVELEKI
Co. burnetii        ISALWGPAHG GANEACLNML RKIG..... .DEKNIGQYI KKAKDKNDPF RLMGFGHRVY .KNYDPRAKV MQKTCYEVLD AVG.RHNEPL FKLAIKLEKI
R. prowazekii       IASLWGPAHG GANEAVINML KEIG..... .SSENIPKYV AKAKDKNDPF RLMGFGHRVY .KSYDPRAAV LKETCKEVLN ELGQLDNNPL LQIAIELEAL
B. subtilis (CitZ)  IGALKGPLHG GANEGVMKML TEIG..... .EVENAEPYI .RAK.LEKKE KIMGFGHRVY .KHGDPRAND LKEMSKRLTN LTGE.SKWYE MSIRIE....
P. furiosus         IGALKGPIHG GAVEEAIKQF MEIG..... .SPEKVEEWF .FKA.LQQKR KIMGAGHRVY .KTYDPRARI FKKYASKL.. ..GD.KKLFE IAERLE....
B. subtilis (CitA)  LGTMKGPLHG GAPSAVTKML EDIG..... .EKEHAEAYL .KEK.LEKGE RLMGFGHRVY .KTKDPRAEA LRQKAEEV.. .AGN.DRDLD LALHVEAEAI
B. coagulans (strain C4) VASLKGHLHG GANEAVMEML QDAQ..... .TVEGFKHLL .HDK.LSKKE KIMGFGHRVY MKKMDPRAAM MKEALKELSA VNGD.DLLLQ MCEAGE....
M. smegmatis        IGALKGSLHG GANEAVMHDM LEIG..... .SAEKAPEWL .HGK.LSRKE KVMGFGHRVY .KNGDSRVPT MKVALEQVAQ VRDG.QRWLD IYNTLE....
T. acidophilum      LAALKGPLHG GAAEAAIAQF DEIK..... .DPAMVEKWF .NDNIINGKK RLMGFGHRVY .KTYDPRAKI FKGIAEKLSS KKPEVHKVYE IATKLEDFGI
```

```
                    401          *                        * *    *  *450                 **                    486
Chicken             NVLLEQGAAA NPWPNVDAHS GVLLQYYGMT E.MNYYTVLF GVSRALGVLA QLI.WSRALG FPLERPKSMS TDGLIAL... ......
Pig                 NVLLEQGKAK NPWPNVDAHS GVLLQYYGMT E.MNYYTVLF GVSRALGVLA QLI.WSRALG FPLERPKSMS TDGLIKLVDS K.....
C. elegans          GILLEQGKAK NPWPNVDSHS GVLLQYFGMT E.MSFYTVLF GVSRALGCLS QLI.WARGMG LPLERPKSHS TDGLIKLALA AKK...
Yeast(glyo)         GVLTEHGKTK NPWPNVDAHS GVLLQYYGLK E.SSFYTVLF GVSRAFGILA QLI.TDRAIG ASIERPKSYS TEKYKELVKN IESKL.
Yeast(mito)         GVLTKHGKTK NPWPNVDSHS GVLLQYYGLT E.ASFYTVLF GVARAIGVLP QLI.IDRAVG APIERPKSFS TEKYKELVKK IESKN.
N. crassa           KVLTEHGKTK NPYPNVDAHS GVLLQHYGLT E.ANYYTVLF GVSRAIGVLP QLI.IDRAVG APIERPKSYS TDKWIEICKK L.....
Te. thermophila     KKLLTYKKIA NPYPNVDCHS GVLLYSLGLT E.YQYYTVVF AVSRALGCMA NLI.WSRAFG LPIERPGSAD LKWFHDKYRE ......
A. anitratum        ALNDPYFVER KLYPNVDFYS GIILKAIGIP ..TEMFTVIF ALARTVGWIS HWLEMHSGP. YKIGRPRQLY TGEVQRDIKR ......
Ps. aeruginosa      ARHDPYFVER NLYPNVDFYS GIILKAIGIP ..TSMFTVIF ALARTVGWIS HWQEMLSGP. YKIGRPRQLY TGHTQRDFTA LKDRG.
E. coli             ALNDPYFIEK KLYPNVDFYS GIILKAMGIP ..SSMFTVIF AMARTVGWIA HWSEMHSDG. MKIARPRQLY IGYEKRDFKS DIKR..
Ac. acetii          ALSDDYFVQR KLYPNVDFYS GIILKAMGIP ..TSMFTVLF AVARTTGWVS QWKEMIEEPG QRISRPRQLY IGAPQRDYVP LAKR..
Co. burnetii        ALEDDYFIEK KLYPNVDFYS GLTLNAIGIP ..SNMFTVIF ALSRTVGWIS HWMEMMSSPD HRLARPRQLY TGETEREVIS LDKRQA
R. prowazekii       ALKDEYFIER KLYPNVDFYS GIIYKAMGIP ..SQMFTVLF AIARTVGWMA QWKEMHEDPE QKISRPRQLY TGYVHREYKC IVERK.
B. subtilis (CitZ)  DIVT...SEK KLPPNVDFYS ASVYHSLGID H..DLFTPIF AVRRMSGWLA HILEQYDN.N .RLIRPRADY TGPDKQKFVP IEERA.
P. furiosus         RLVEEYLSKK GISINVDYWS GLVFYGMKIP I..ELYTTIF AMGRIAGWTA HLAEYVSH.N .RIIRPRLQY VGEIGKKYLP IELRR.
B. subtilis (CitA)  RLLEIYKPGR KLYTNVEFYA AAVMRAIDFD D..ELFTPTF SASRMVGWCA HVLEQAEN.N .MIFRPSAQY TGAIPEEVLS ......
B. coagulans (strain C4) ...QIMREEK GLFPNLDYYA APVYWKLGIP I..PLYTPIF FSSRTVGLCA HVMEQHEN.N .RIFRPRVLY TGARNLRVED ......
M. smegmatis        ...SAMFAAT RIKPNLDF.P TGAYYLMDFP I..ESFTPLF VMSRITGWTA HIMEQAAS.N .ALIRPLSEY SGQPQRSLV. ......
T. acidophilum      KAF....GSK GIYPNTDYFS GIVYMSIGFP LRNNIYTALF ALSRVTGWQA HFIEYVEEQQ .RLIRPRAVY VGPAERKYVP IAERK.
```

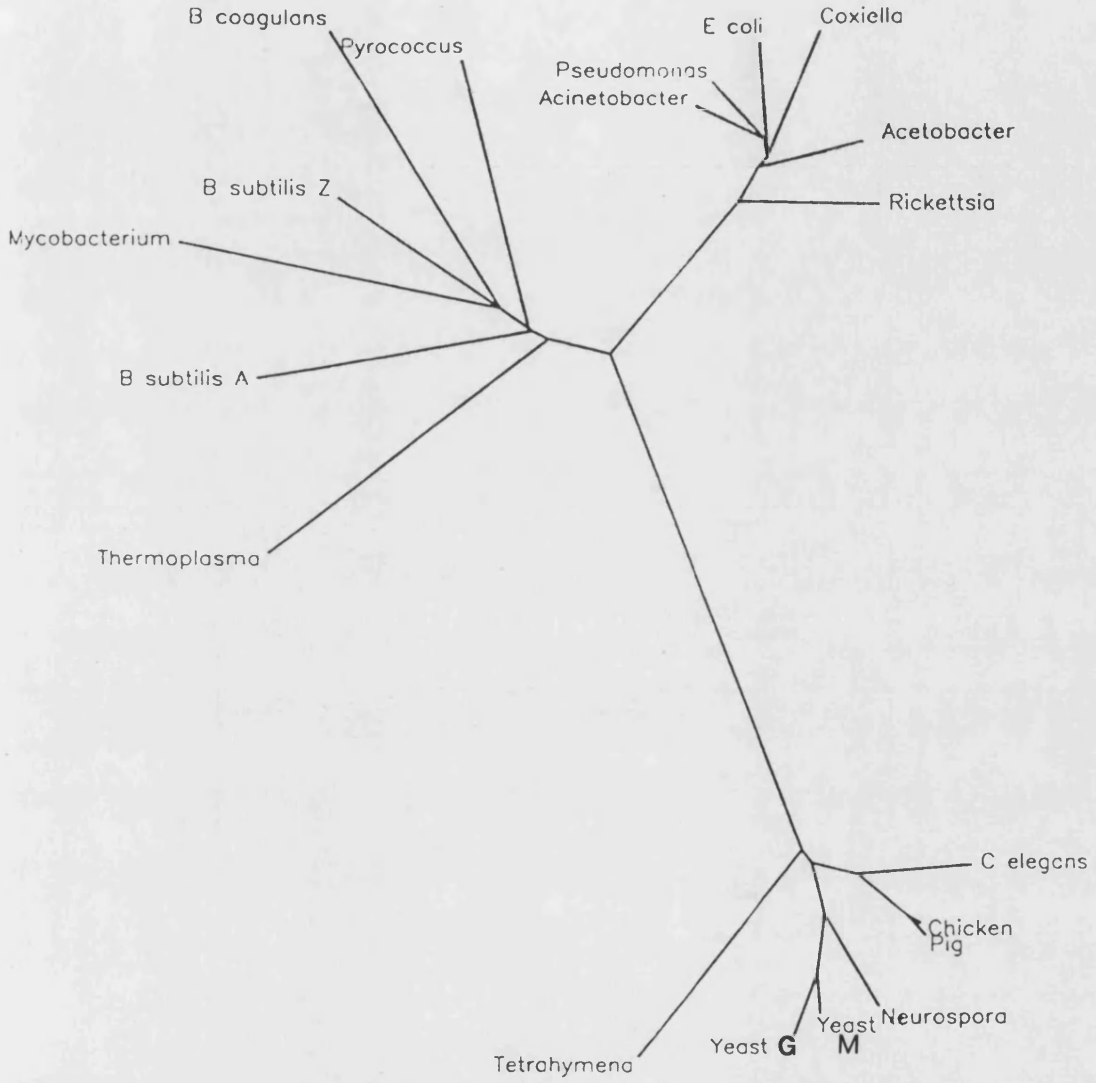


Figure 7.3 Phylogenetic tree generated from known citrate synthase sequences using PHYLIP Version 3.5c [Felsenstein, 1993]. A distance matrix was generated from a multiple sequence alignment using the program PROTDIST. From this, a phylogenetic tree was created using the FITCH program which does not assume a molecular clock.

The phylogenetic tree, generated by this analysis, does not agree with the global tree topology suggested by Woese *et al.* [1990] (Figure 1.6) in that the sequences of the archaeal species do not form a monophyletic group. The unexpected tree topology generated by this analysis may be explained by the assessment of the degree of sequence identity between the citrate synthase sequences of each cluster (Table 7.2). From this, it can be seen that the identity between sequences within the eukaryal cluster is high (49.8 % - 91.9 %). The eukaryal sequences also cluster in an expected manner; the chicken, pig and *C. elegans* sequences cluster together, as do the citrate synthases of fungal organisms, *N. crassa* and yeast. The primative protozoan, *Tetrahymena thermophila* branches off at the base of this cluster. The citrate synthase sequences of Gram-negative Bacteria also exhibit a high degree of identity (57.0 % - 75.8 %). In contrast, the identity between eukaryal and Gram-negative bacterial sequences is low (25.2 % - 30.7 %).

The identity between the citrate synthases sequences of Gram-positive bacterial and archaeal organisms form a more disparate group, exhibiting only 32.2 % - 46.0 % identity between members. This is only marginally higher than the identity shown between sequences of this cluster with the sequences of the other two clusters (22.4 % - 30.5 % and 31.2 % - 38.8 % identity with the citrate synthases of Eukarya and Gram-negative Bacteria, respectively). Bootstrap resampling also demonstrates disparity within the Archaeal/Gram-positive Bacteria cluster. Of the 100 resampled trees, 100 % clustered the eukaryal and Gram-negative bacterial sequences as two separate groups yet only 75 % generated the Archaea/Gram-positive Bacteria cluster indicated in Figure 7.3. In addition, the two archaeal sequences branch off at the base of the cluster which may indicate that on the determination of more sequences, these may form a discrete cluster.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 Chicken		91.9	68.6	56.7	59.3	60.4	54.0	27.3	28.5	26.2	30.1	29.0	27.1	28.9	30.5	24.3	26.5	27.5	22.4
2 Pig			71.4	58.9	61.7	62.8	55.7	29.8	26.2	26.3	29.9	28.0	27.0	29.4	30.7	24.7	28.6	27.6	22.5
3 *C. elegans*				56.4	58.9	57.4	50.0	29.5	27.9	27.5	31.9	28.8	27.5	27.8	30.9	24.0	28.3	26.1	25.4
4 Yeast (glyoxysomal)					81.8	64.9	46.9	26.0	26.5	25.2	28.7	26.0	25.5	25.6	27.3	26.6	28.5	24.8	24.7
5 Yeast (mitochondrial)						69.0	50.7	27.2	26.2	26.7	28.3	26.2	27.9	27.1	26.7	26.9	27.1	25.3	26.3
6 *N. crassa*							49.8	26.2	27.7	27.1	30.2	28.3	29.2	26.1	28.8	29.1	26.7	24.2	24.5
7 *Te. thermophila*								29.5	27.5	26.6	30.7	29.0	28.1	28.0	28.3	25.1	25.4	24.7	26.6
8 *A. anitratum*									75.8	67.3	68.7	65.2	60.7	36.2	35.7	34.7	34.0	35.9	31.2
9 *Ps. aeruginosa*										68.3	67.6	66.7	60.2	36.5	34.0	32.9	32.9	31.5	31.6
10 *E. coli*											65.0	61.2	60.0	35.2	35.9	31.9	33.2	30.0	32.4
11 *Ac. acetii*												60.7	57.3	38.8	34.9	34.5	32.9	34.8	33.9
12 *Co. burnetii*													57.0	37.6	35.9	32.9	32.3	31.7	35.4
13 *R. prowazekii*														36.3	36.1	35.1	31.3	33.3	34.0
14 *B. subtilis* (CitZ)															46.0	41.9	42.7	43.5	36.2
15 *P. furiosus*																37.3	37.5	41.3	42.1
16 *B. subtilis* (CitA)																	36.6	40.4	32.2
17 *B. coagulans* (strain C4)																		35.1	31.3
18 *M. smegmatis*																			32.6
19 *T. acidophilum*																			

Table 7.2 Percentage sequence identites between each pair of known citrate synthase sequences as determined using BESTFIT (University of Wisconsin GCG)

It must be noted, that the recent demonstration that organisms such as *B. subtilis*, *Ps. aeruginosa* and *E. coli* possess two citrate synthase sequences necessitates a cautious approach to this protein-derived phylogenetic analysis. It must be ensured that the sequences of equivalent enzymes are being compared. For example, the second citrate synthase of *Ps. aeruginosa* and *E. coli* is dimeric and amino-terminal sequencing has suggested that the amino-terminus may be recessed, in common with the enzyme from Gram-positive bacterial and archaeal sources.

7.2.4 Putative thermostabilising features of the *P. furiosus* citrate synthase

The aim of this discussion section is to assimilate the features inferred by the compositional analysis (Table 7.3), structural alignment (Figure 7.4) and homology modelling of the *P. furiosus* citrate synthase sequence with respect to protein thermostability. As the chosen approach to this investigation is one of comparative enzymology, central to this discussion will be the putative thermostabilising features exhibited by the *T. acidophilum* enzyme, as deduced by a comparison of the crystal structure of this enzyme with that of the pig heart citrate synthase [Russell *et al.*, 1994]. Attempts will then be made to observe if these features extend to the *P. furiosus* enzyme, or if alternative thermostabilising mechanisms are operating. Following this comparison, a number of additional features observed in the *P. furiosus* citrate synthase will be described. Although it is not possible to gauge the effects of these features at present, they may be targetted for detailed analysis upon the elucidation of the crystal structure.

	G	A	V	I	L	P	M	C	S	T	Y	W	F	D	E	Q	N	K	R	H
Chicken	9.0	10.4	5.6	3.9	12.7	5.1	3.2	0.7	6.9	5.1	4.2	2.1	3.0	3.9	4.2	3.9	3.9	4.2	4.4	2.8
Pig	7.6	7.6	6.4	4.3	12.1	5.0	3.4	0.9	6.6	5.3	4.3	2.1	2.7	4.8	5.5	3.9	4.1	5.7	4.3	3.2
Yeast (g)	7.3	7.0	6.8	4.8	11.6	4.8	1.6	0	8.6	3.9	4.5	1.4	3.6	6.4	6.8	3.6	3.2	7.3	3.9	3.0
Yeast (m)	7.7	8.4	6.1	5.4	11.8	5.2	0.7	0	6.3	5.2	4.3	1.4	3.6	4.3	8.8	2.9	2.5	8.4	4.1	2.7
Te. thermophila	7.0	6.3	6.6	5.7	10.4	4.5	2.7	1.6	5.9	3.2	5.2	1.6	3.2	7.3	4.8	4.8	2.9	8.6	3.4	4.3
A. anitratum	7.1	9.9	6.6	6.9	8.7	4.7	3.3	0.9	4.0	5.2	4.0	0.9	4.7	6.1	6.1	2.4	4.0	5.4	5.2	3.5
Ps. aeruginosa	7.5	8.9	5.6	6.1	8.4	5.6	4.2	1.4	4.9	5.6	3.5	0.7	4.9	5.9	5.4	3.0	4.0	6.1	4.4	3.7
E. coli	6.6	8.7	4.5	7.0	8.2	4.5	4.2	1.6	5.9	6.1	5.8	0.7	5.4	6.3	6.1	1.9	4.0	5.4	5.6	3.5
Ac. acetii	7.1	11.3	5.3	7.1	9.4	5.5	3.0	0.9	4.6	5.3	3.9	0.9	4.4	5.3	4.8	3.4	5.1	5.5	5.1	2.1
Co. burnetii	6.5	9.1	4.7	6.3	8.4	4.9	4.0	1.2	5.6	5.1	5.4	0.7	4.9	5.6	6.3	2.1	4.2	7.0	4.9	3.3
R. prowazekii	5.5	8.7	4.1	8.3	8.5	4.6	3.2	1.6	6.9	4.6	5.7	0.9	5.1	5.1	6.0	2.5	5.5	7.1	3.4	2.5
B. subtilis (CitZ)	6.3	11.0	5.5	5.8	9.3	3.8	3.0	0.5	5.8	6.3	4.4	0.5	3.0	3.8	10.4	2.2	3.0	6.0	5.5	3.8
P. furiosus	7.7	7.4	5.6	9.0	9.8	4.5	2.4	0.3	5.1	4.0	6.6	1.9	2.7	2.9	10.4	1.3	2.9	8.8	4.5	2.1
B. subtilis (CitA)	5.9	8.9	5.4	7.3	10.2	4.9	2.2	0.3	5.9	5.4	4.0	0.8	3.0	5.1	10.0	1.6	4.3	6.2	5.7	3.0
B. coagulans	5.9	7.0	6.2	5.6	13.4	4.6	3.2	0.5	5.9	5.4	4.0	0.3	3.2	4.6	8.6	3.2	4.3	6.5	3.8	3.8
M. smegmatis	6.2	12.4	8.1	4.3	10.5	5.9	3.8	0.8	7.5	6.7	3.5	1.1	3.8	4.6	6.7	5.6	3.0	3.2	8.1	3.0
T. acidophilum	6.8	11.5	5.7	8.1	6.3	4.4	3.1	0	4.9	5.5	5.2	1.0	3.9	4.9	7.8	2.6	3.9	7.8	4.9	1.6

Table 7.3 Comparison of the percentage amino acid content of citrate synthases sequences

The citrate synthase sequences of *N. crassa* and *C. elegans* were omitted from this analysis as their primary sequences were derived from the nucleotide sequence and thus included the mitochondrial leader sequence. Yeast (g) and Yeast (m) represent the glyoxysomal and mitochondrial form of the enzyme, respectively.

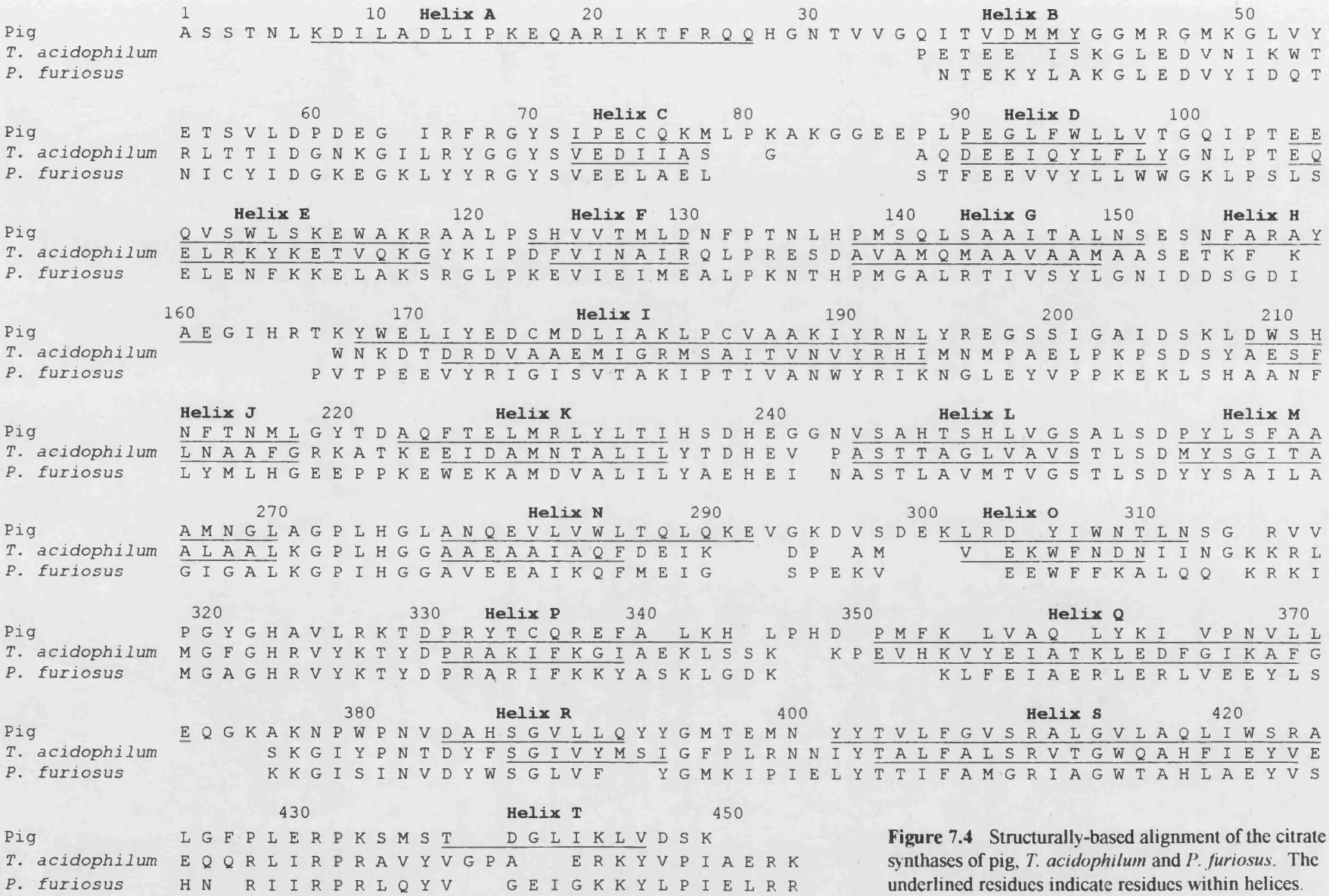


Figure 7.4 Structurally-based alignment of the citrate synthases of pig, *T. acidophilum* and *P. furiosus*. The underlined residues indicate residues within helices.

The structure of the pig and *T. acidophilum* citrate synthase is described in Sections 1.1.2 and 1.1.4, respectively. To aid this discussion, Figure 7.5 illustrates the structure of the *T. acidophilum* monomer, resolved to 2.5Å [Russell *et al.*, 1994]. To recapitulate, each subunit consists of a large and a small domain, between which is located the active site. The dimer is generated by a two-fold symmetry operation such that α-helices F, G, L and M form antiparallel α-helical pairs with helices F', G', L' and M' (respectively); these are orientated perpendicularly to the axis of symmetry. Residues from each subunit contribute to each of the two active sites per dimer. The *P. furiosus* enzyme was homology-modelled to the *T. acidophilum* citrate synthase monomer. Therefore, prior to a discussion of the features exhibited by the *P. furiosus* enzyme it is necessary to consider the quality and limitations of the homology-modelled structure.

The quality of the homology-modelled citrate synthase

The *P. furiosus* primary sequence was homology-modelled to the crystal structure of the *T. acidophilum* enzyme, resolved to 2.5Å. The quality of the model was demonstrated by Ramachandran plots which did not indicate any residues in disallowed conformations, and by a 3D-1D profile (the output of the ENVIRONMENTS and VERIFY programs) which demonstrated that each residue was in an acceptible environment. In addition, the stereochemistry of the active site (Figure 7.6) was identical to that of the *T. acidophilum* citrate synthase.

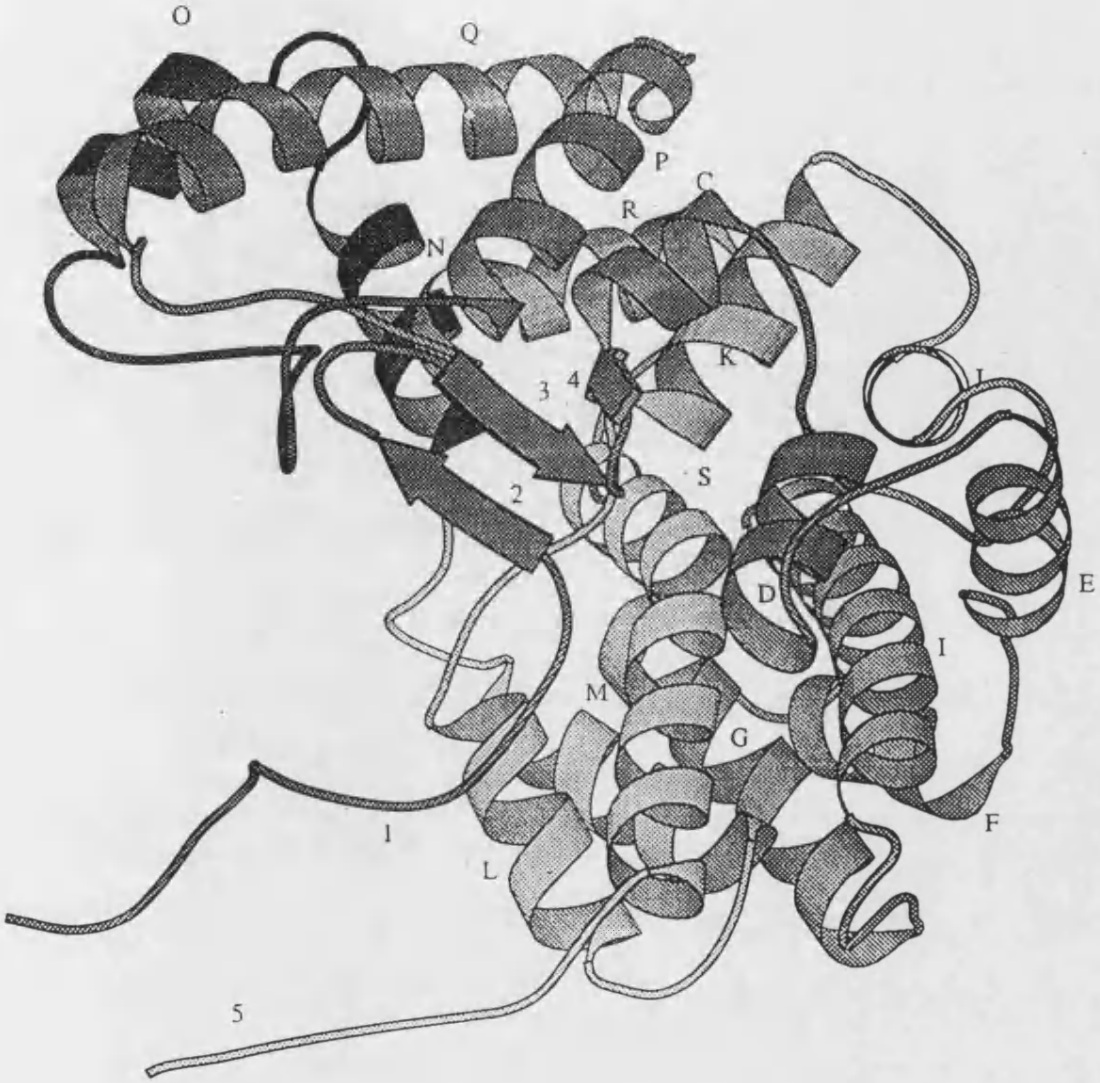


Figure 7.5 Schematic representation of the *T. acidophilum* citrate synthase monomer Strands of β-sheet are denoted by numbers 2 - 4 whereas each α-helix is represented by a letter. The notation of the α-helices corresponds directly to the pig heart enzyme in that helix A and helix H are missing. Due to diffuse electron density, helices B and T are also not defined in the *T. acidophilum* enzyme; these regions are denoted by the numbers 1 and 5. [reproduced from Russell *et al.*, 1994]

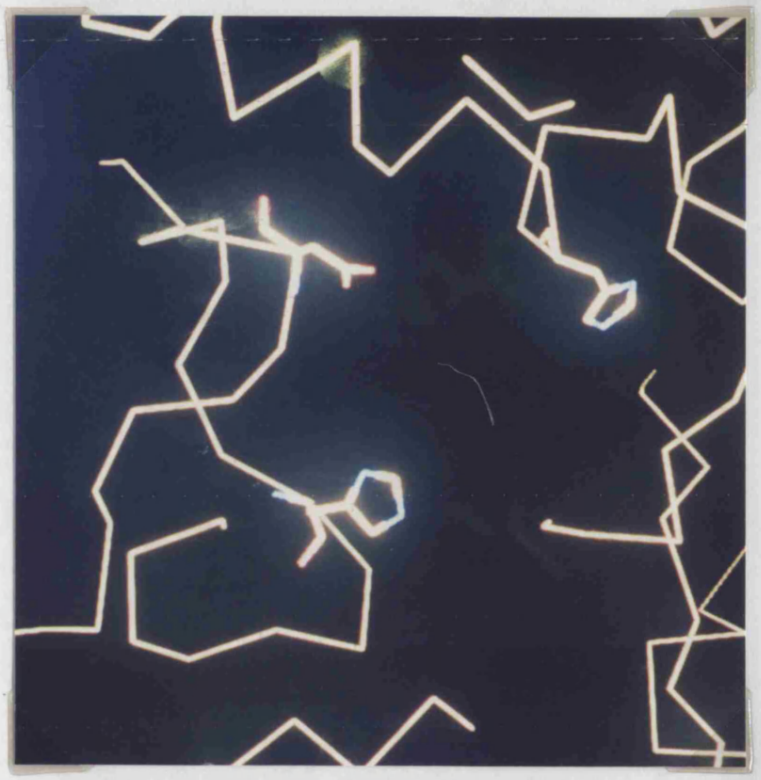


Figure 7.6 The active site of the *P. furiosus* citrate synthase, homology-modelled to the *T. acidophilum* crystal structure at 2.5 Å resolution. The residues illustrated, His224, His263 and Asp313, are equivalent to those in the pig heart enzyme that have been implicated in the reaction mechanism (Section 1.1.3).

Prior to discussing the putative thermostabilising features of this enzyme, it is necessary to recommend caution when interpreting the model for two reasons. First, the sequence was modelled to a low resolution crystal structure and second, the absence of crystallographic data necessitated the manual selection of side chain rotamers for each substituted residue or loop rearrangement. Thus, most surface sidechains are extended into the solvent whereas internal sidechains with the most favourable stereochemistry were selected by eye. This limits the information provided by the modelling exercise because specific interactions cannot be assessed in detail. However, the homology-modelling of the *P. furiosus* citrate synthase sequence is a powerful means of visualising the primary sequence, especially when the two sequences show a high percentage identity, as in this case. The secondary structure of the *P. furiosus* citrate synthase can be depicted as helical wheels (Figure 7.7). These demonstrate the degree of solvent exposure of residues in these structural motifs thereby providing a convenient way of demonstrating any patches of surface charge, hydrophobicity or α-helix amphipathicity.

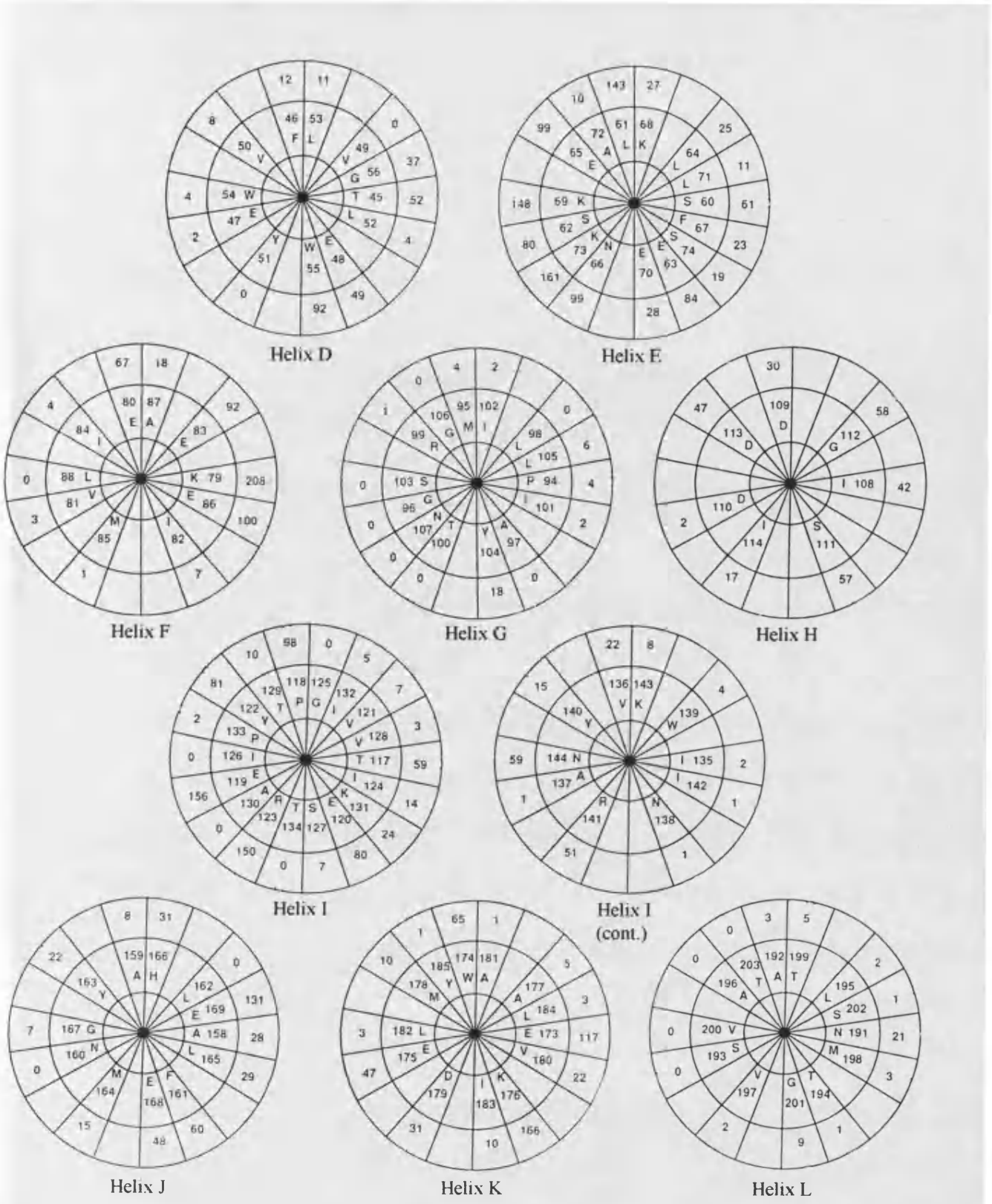
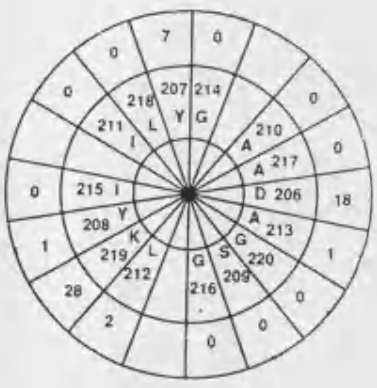
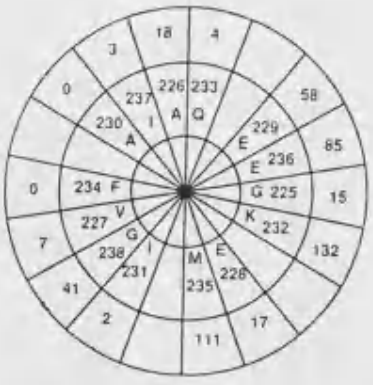


Figure 7.7 Helical wheels generated from the homology modelled *P. furiosus* citrate synthase

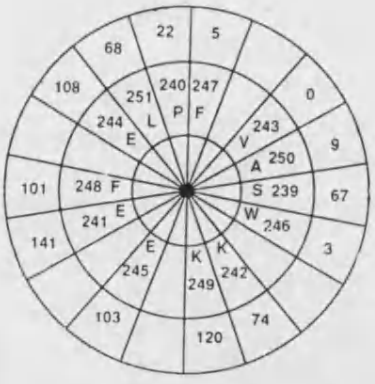
The inner numbers represent the amino acid primary sequence, the outer numbers represent the solvent-accessible surface area in $Å^2$



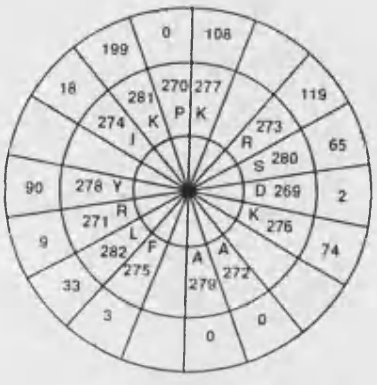
Helix M



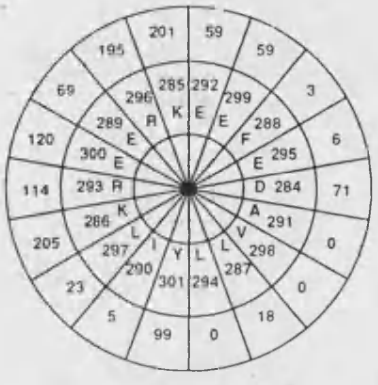
Helix N



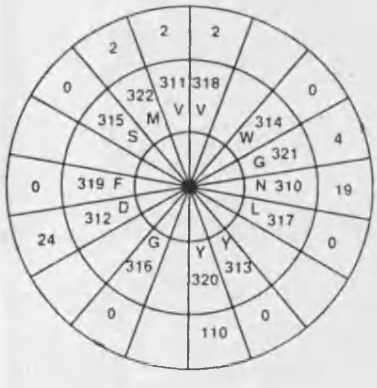
Helix O



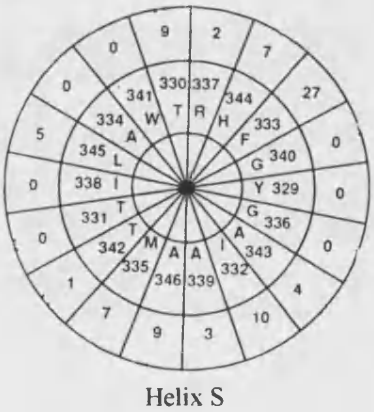
Helix P



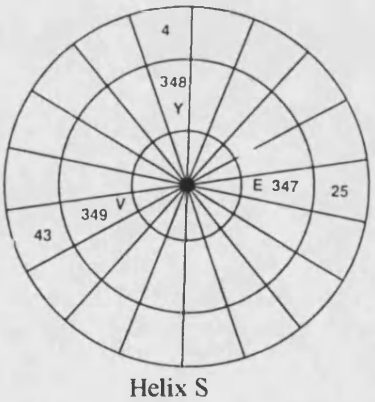
Helix Q



Helix R



Helix S



Helix S
(cont.)

Figure 7.7 Helical wheels generated from the homology-modelled *P. furiosus* citrate synthase (continued)

7.2.4.1 Comparison of the pig, *T. acidophilum* and *P. furiosus* citrate synthase

The comparison of the *T. acidophilum* and pig heart citrate synthase crystal structures [Russell *et al.*, 1994] generated a number of putative thermostabilising features namely, a recessed amino-terminus, shorter loops between α-helices, smaller internal cavities and increased stability at the subunit interface by residue substitutions to Ala in the *T. acidophilum* enzyme. The moderately thermostable enzyme also exhibited increased aromatic interactions in the small subunit of the citrate synthase monomer.

The recessed amino terminus

Both the *T. acidophilum* and the *P. furiosus* citrate synthases are shorter than the pig heart citrate synthase. As shown in the multiple sequence alignment (Figure 7.2), this is due in part, to the absence of approximately 30 residues at the amino-terminus. The crystal structure of the pig heart enzyme demonstrates that these residues comprise helix A which rests in a groove on the surface of the second subunit, and the long loop region that connects this α-helix to helix B (Figure 1.4). As discussed in Section 3.4.1, the elucidation of the citrate synthase sequences of *B. subtilis* and the confirmation of the translation start site of the *M. smegmatis* citrate synthase indicates that the feature of a recessed amino-terminus is shared by the enzyme from Gram-positive bacterial sources. The thermostability of the citrate synthases of these Gram-positive Bacteria has not been measured. If they do not exhibit thermostability then the possession of a recessed amino-terminus is likely to be a phylogenetically-derived characteristic.

Shorter loop regions

The *T. acidophilum* citrate synthase exhibits shorter loop regions between α-helices than the enzyme from pig heart. The subsequent decrease in flexibility generated by such an adaptation is likely to stabilise the folded conformation. It is therefore not surprising that the *P. furiosus* citrate synthase exhibits shorter loops. In addition, the *P. furiosus* enzyme also exhibits one loop region, between helices P and Q (residues 248-252 of the structural alignment (Figure 7.4)) which is even shorter than that of the *T. acidophilum* enzyme. This may be contributing to the stability of the small domain by enforcing a tight, inflexible turn enhancing the rigidity of this domain. From Figure 7.4, it can also be seen that a Gly residue is present at the carboxyl-end of helix P. The dihedral angles that this residue can adopt may be essential in effecting the tight turn.

The unusual feature of two adjacent Pro residues can be seen in the loop regions between helices I and J and J and K (positions 201-202 and 221-222 of the structural alignment (Figure 7.4)). Only one other citrate synthase, that of *M. smegmatis*, has two adjacent Pro residues; the multiple sequence alignment demonstrates that these are conserved with the *P. furiosus* enzyme between helices I and J (position 229-230 of the multiple sequence alignment (Figure 7.2)). The restriction in conformation, generated by such a linkage, may be influential in increasing the stability of the folded conformation.

Internal cavities

The contribution of internal packing to the stability of the folded conformation has been illustrated in a number of enzymes by the effective 'truncation' of the sidechains lining the cavity by site directed mutagenesis. As described in detail in Section 1.4.2, a correlation was observed between the ability of the neighbouring residues to collapse into the cavity, minimising its size, with the effect of the mutation on stability [Shortle *et al.*, 1990]. This highlighted the importance to stability of a tightly-packed internal core that maximises favourable van der Waals interactions. In agreement with this study, fewer cavities were observed in the *T. acidophilum* citrate synthase crystal structure than in the pig enzyme, and those remaining were smaller in size (comprising 0.25 % and 0.6 % of the total volume of the *T. acidophilum* and pig enzyme, respectively).

Unfortunately, due to the limitations of the homology model, it was not valid to measure the size of internal cavities in the *P. furiosus* enzyme; this analysis requires the elucidation of the crystal structure of the enzyme. However, it is interesting to note that the amino acid composition comparison (Figure 7.3) demonstrates that the *P. furiosus* enzyme exhibits a high percentage of Leu, Ile, Tyr and Trp residues, all of which have bulky, hydrophobic sidechains and may contribute to the compact enzyme core. This would suggest an extension of the trend observed in the *T. acidophilum* enzyme.

The subunit interface

As mentioned previously, the subunit interface is composed of α-helices F, G, L and M which form antiparallel helical pairs with α-helices F', G', L' and M' of the second subunit. As can be seen from the structural alignment, helix G (residues 138-151 of Figure 7.4) of the *T. acidophilum* enzyme differs from the equivalent pig heart helix in that it is rich in Ala residues. These are orientated towards the helix G' to allow hydrophobic interactions. As Ala is considered to have the highest helical propensity [Serrano *et al.*, 1992a], this feature is in agreement with previous studies detailed in Section 1.4.2. However, this feature is not extended to the *P. furiosus* helix G which exhibits greater identity to the pig enzyme in this region. In addition, the overall percentage Ala content of the *P. furiosus* citrate synthase (Figure 7.3) is lower than the *T. acidophilum* enzyme; this also may indicate that an alternative mechanism of stabilisation is operating.

Recent proposals [reviewed by Dill, 1990] have suggested that at extreme temperatures (> 100°C), the contribution to stability of the hydrophobic effect is less than at slightly lower temperatures (~80°C). This would explain the observation of alternative methods of stabilisation in the moderately and extremely thermophilic citrate synthase.

Aromatic interactions

The final feature highlighted by the comparison of the *T. acidophilum* and pig crystal structures was the demonstration of increased aromatic interactions in the small domain of the *T. acidophilum* monomer. Although the homology model suggests that aromatic interactions occur in the small domain of the *P. furiosus* citrate synthase subunit (Figure 7.8), it is not possible to assess this feature quantitatively due to the limitations of the homology model; However, the percentage of aromatic residues in the *P. furiosus* citrate synthase is one of the highest of all citrate synthases, second only to that of the Gram-negative Bacterium, *R. prowazekii* (Table 7.3). Moreover, the *P. furiosus* enzyme exhibits the highest Tyr content and second highest Trp content of all known citrate synthases, despite Trp being encoded by the single guanosine-rich codon, TGG. Helix D exhibits two adjacent Trp residues at positions 98-99 of the structural alignment (Figure 7.4) followed by a Gly which may be acting as a carboxyl-capping residue. The homology model of the *P. furiosus* citrate synthase inferred that these Trp residues may be interleaved by a non-local Tyr residue, thus promoting favourable van der Waals interactions (Figure 7.9). Thus, the trend observed in the *T. acidophilum* enzyme may be extended in the *P. furiosus* enzyme, although elucidation of the crystal structure is necessary for accurate assessment of these features.

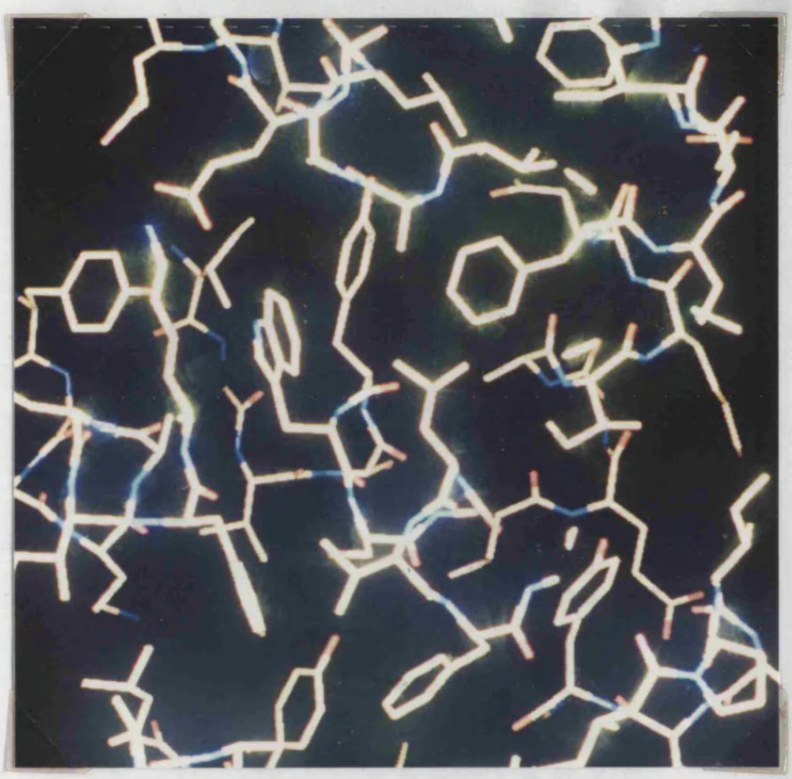


Figure 7.8 An internal region of the small domain of the *P. furiosus* citrate synthase, homology-modelled to the *T. acidophilum* crystal structure at 2.5 Å resolution, showing aromatic packing.

Figure 7.9 An internal region of the large domain of the *P. furiosus* citrate synthase, homology-modelled to the *T. acidophilum* crystal structure at 2.5 Å resolution, showing the two adacent Trp residues of helix D interleaved by a Tyr residue to promote favourable van der Waals interactions.

7.2.4.2 Other features of the *P. furiosus* citrate synthase

Avoidance of thermolabile residues

As well as the features outlined above, the *P. furiosus* enzyme also exhibits a decrease in the number of residues that are known to be susceptible to chemical modification at high temperatures. For example, Cys residues are susceptible to oxidation at high temperatures [Zale and Klibanov, 1986]. The Cys content of the *P. furiosus* enzyme is limited to one, located in the three strands of β-sheet near the amino-terminus of the enzyme. This represents a reduction in Cys residues over the citrate synthases of most other sources and thus, may suggest a discrimination against this residue as a mechanism of thermostabilisation. However, it must be noted that the citrate synthases of yeast and *T. acidophilum* do not have any Cys residues.

The percentage of Asn and Gln residues is the lowest found in all citrate synthases. This discrimination is likely to be a mechanism by which the enzyme avoids chemical modification by deamidation, as described in Section 1.4.3. Asp is also discriminated against in the *P. furiosus* enzyme (Figure 7.3). This may be a preventative mechanism by which the enzyme avoids chemical cleavage of the polypeptide chain. One Asp-Gly linkage, which is the most thermolabile peptide bond, is present in the amino-terminal portion of the *P. furiosus* enzyme (position 59-60 of the structural alignment (Figure 7.4)). These two residues are also conserved in the *T. acidophilum* enzyme and thus, may be important to the overall structure or mechanism of the enzyme.

The charged residues of the P. furiosus citrate synthase

The composition of charged residues, Asp, Glu, Lys, Arg and His is also unexpected with regard to previous studies. The composition of the *P. furiosus* enolase, determined recently by Peak *et al.*, [1994], suggested a correlation between enzyme thermostability and the ratio of hydrophobic to hydrophilic residues, as calculated by the sum of the number of Ala, Val, Ile, Leu and Phe residues divided by the sum of the number of Asp, Glu, Lys and Arg. This was not found to be the case with the *P. furiosus* citrate synthase as the enzyme had one of the lowest ratios of all known citrate synthases.

The Lys and Arg content of both the *P. furiosus* and the *T. acidophilum* citrate synthases do not follow the trend, suggested by previous studies [Menéndez-Arias and Argos, 1989; Qaw and Brewer, 1986], that substitution of Lys by Arg at solvent-exposed positions increases the thermostability of an enzyme. In the *P. furiosus* citrate synthase, the ratio Arg/(Arg+Lys) is the second lowest of all known citrate synthases, illustrating that the reverse trend is operating in this enzyme (helices E and Q of Figure 7.7).

The surface charge can be illustrated on the homology model, shown in Figure 7.10. Interestingly, the surface charge generated by the intersubunit helix F appears to be orientated in close proximity to that of the antiparallel counterpart of this helix, helix F' of the second subunit. It may be possible that repulsion of these charged regions (which would favour the unfolding of the enzyme) may be prevented by positively-charged ions in the cellular environment. As mentioned earlier, the recombinant enzyme was activated by both NaCl and KCl, thereby inferring that charged ions do interact with the enzyme molecule.

Enzymes from halophilic sources also exhibit patches of surface charge; these are thought to co-ordinate a shell of water molecules around the enzyme to maintain the conformation within the high-salt environment [Lanyi, 1974; Zaccai and Eisenberg, 1990]. Removal of the salt causes the charged patches to repel and the enzyme to unfold. The surface charge may also be involved in the thermostabilisation of the enzyme via extrinsic factors in the cell cytoplasm, as discussed in Section 6.4.3.

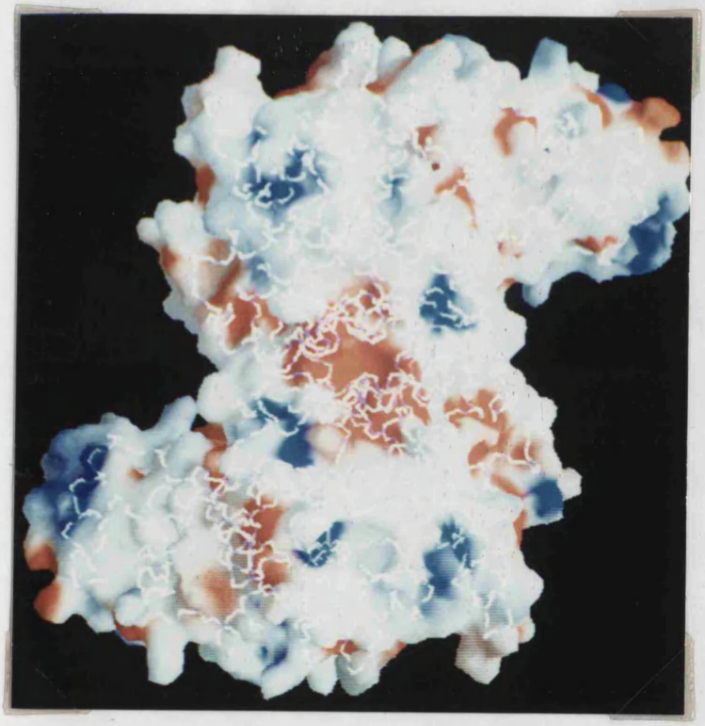


Figure 7.10 Space-filling model of the homology-modelled *P. furiosus* citrate synthase dimer, looking down the two-fold axis of symmetry. The surface is coloured according to the electrostatic potential with regions of positive charge (blue) and negative charge (red). The central region of negative charge derives from the two antiparallel intersubunit helices F and F'.

Conclusions

The above discussions highlight a number of putative thermostabilising features of the *P. furiosus* citrate synthase. They also illustrate the dependence of detailed observations on the elucidation of the crystal structure of this enzyme. Hence, the following chapter details the progress made in the final weeks of this project towards this ultimate aim.

CHAPTER 8

PRELIMINARY CRYSTALLOGRAPHIC ANALYSIS OF THE *P. FURIOSUS* CITRATE SYNTHASE

8.0 INTRODUCTION

The previous chapter has described a number of features that appear novel to the *P. furiosus* citrate synthase, such as a shorter loop region, the predominance of Lys and Glu residues at solvent-exposed positions, adjacent Pro residues in loop regions and a high content of Tyr and Trp residues. The homology model of the *P. furiosus* citrate synthase, constructed using the crystal structure of the *T. acidophilum* enzyme has been essential in visualising the primary sequence but it has certain limitations, as detailed in Section 7.2.4. Therefore, to enable specific interactions to be investigated and to allow measurement of features such as the size of internal cavities and aromatic interactions, an accurate illustration of the structure is required, as determined by X-ray crystallography.

An ordered array of protein molecules, in the form of a protein crystal, will diffract X-rays of wavelength similar to the distances between atoms within a molecule (1.5Å). The basic repeating unit that generates the crystal by translation in three dimensions without rotating, is the unit cell. The molecules within the unit cell are related to each other by rotational and translational symmetry operations; the individual unit that is repeated in these operations is the asymmetric unit. There are 230 possible crystal lattices, known as space groups. Proteins do not crystallise in all space groups as they cannot exhibit reflectional symmetry due to the chirality of amino acids.

This chapter describes the crystallisation of the *P. furiosus* citrate synthase and the preliminary crystallographic analyses to determine the unit cell dimensions and the space group.

8.1 MATERIALS

Reagents for the Hampton Crystal Screen™ were obtained from Hampton Research, Riverside, USA. Quartz capillary tubes of diameter 0.5 mm were obtained from GLAS, Berlin, Germany

8.2 METHODS

8.2.1 Crystallisation of the *P. furiosus* citrate synthase

The recombinant citrate synthase of *P. furiosus* was purified from the *E. coli* JM105 clone, PfCS(EXP5), as described in Section 6.2.4. Using Centriprep concentrators (Amicon), the purified enzyme was buffer-exchanged to 20 mM Tris-HCl pH 8.0 containing 25 mM NaCl, and concentrated to 5 mg protein/ml. Crystals were then prepared by the hanging drop, vapour diffusion method. A 0.7 ml volume of each reagent of the Hampton Research Crystal Screen was placed in each well of a culture dish. The edges of the well were then smeared with paraffin wax. Upon a silane-treated cover slip, a 2 µl volume of the protein solution was mixed with a 2 µl volume of the screen reagent and the coverslip inverted over the appropriate well. An air-tight seal between the coverslip and the well was ensured by the paraffin wax.

8.2.2 Preliminary crystallographic analysis of the *P. furiosus* citrate synthase

The crystal was mounted in a 0.5 mm quartz capillary tube with a small amount of the mother liquor, positioned in a beam of X-rays produced by a Siemens rotating anode and the diffraction pattern recorded on a Siemens Area Detector (G. L. Taylor, University of Bath). A total of 360 frames of data were collected; between each frame the crystal was rotated by ¼°. The data were then processed using the computer software package XDS [Kabsch, 1988].

8.2.3 Preliminary crystallographic analysis of the *P. furiosus* citrate synthase under Cryostream

The crystal was incubated for 2-3 min at room temperature in a drop of the mother liquor supplemented with glycerol to 20 % (v/v). The crystal was then mounted in a fine mohair loop, positioned in the path of the X-ray beam and frozen immediately in a stream of gaseous nitrogen at 100°K produced by a Cryostream system (Oxford Cryosystems, Oxford, UK.) (G.L. Taylor, University of Bath). Under these conditions, 600 frames of data were collected; between each frame the crystal was rotated by ¼°. The data were then processed using the computer software package XDS [Kabsch, 1988].

8.3 RESULTS

8.3.1 Crystallisation of the *P. furiosus* citrate synthase

The *P. furiosus* citrate synthase crystallised in 22 of the 50 conditions of the Hampton Research Crystal Screen. The crystals were observed after approximately 1 h and were tetragonal bipyramidal in shape, as illustrated in Figure 8.1. Table 8.1 lists the buffer conditions that produced the most defined crystals.

Buffer conditions that generated small tetragonal bipyramidal crystals
0.4 M ammonium dihydrogen phosphate
0.1 M sodium acetate pH 4.6, 2.0 M sodium formate
0.1 M sodium Hepes pH 7.5, 0.8 M potassium phosphate monobasic, 0.8 M sodium phosphate monobasic
0.1 M sodium cacodylate pH 6.5, 1.4 M sodium acetate trihydrate
0.1 M Tris-HCl pH 8.5, 2.0 M ammonium dihydrogen phosphate

Buffer conditions that generated large tetragonal bipyramidal crystals
0.1 M sodium citrate pH 6.5, 1.0 M ammonium dihydrogen phosphate
0.1 M sodium Hepes pH 7.5, 0.8 M potassium-sodium tartrate tetrahydrate
0.1 M sodium Hepes pH 7.5, 1.5 M lithium sulphate monohydrate
2.0 M ammonium sulphate

Table 8.1 Buffer conditions in which the *P. furiosus* citrate synthase crystallised successfully.

Figure 8.1 Crystals of the *P. furiosus* citrate synthase

The large tetragonal bipyramidal crystals were generated in 0.1 M sodium Hepes pH 7.5 containing 1.5 M lithium sulphate monohydrate.

8.3.2 Preliminary crystallographic analysis of the *P. furiosus* citrate synthase

A total of 360 frames of crystallographic data were obtained from a large tetragonal bipyramid crystal, generated in 0.1 M sodium Hepes containing 1.5 M lithium sulphate monohydrate. Due to deterioration of the crystal within the X-ray beam, only 240 frames of data were processed using the XDS software [Kabsch, 1988]. This computed that the data were 63.9 % complete to 4.5 Å with an overall R-merge value of 8.1 %. The R-merge value indicates the quality of the data, and ideally should be less than 10 %. The unit cell dimensions were computed as 99.6 Å x 99.6 Å x 243.4 Å, thereby indicating a tetragonal crystal lattice. Analysis of the systematic absences in the data indicated that the space group was likely to be $P4_12_12$ or $P4_32_12$. A similar analysis of a small tetragonal bipyramidal crystal, generated in 0.4 M ammonium dihydrogen phosphate, demonstrated that the smaller crystal had the same unit cell dimensions and space group.

8.3.3 Preliminary crystallographic analysis of the *P. furiosus* citrate synthase under Cryostream

A large tetragonal bipyramid crystal generated in 0.1 M sodium Hepes pH 7.5 containing 0.8 M sodium-potassium tartrate tetrahydrate was mounted in a mohair loop in the mother liquor, supplemented with glycerol to a concentration of 20 % (v/v). Processing the data using the XDS software [Kabsch, 1988] indicated that the data were 79.3 % complete to 4 Å with an overall R-merge value of 10.6 %. Detailed analysis of the R-merge values indicated that at high resolution (3.0 - 3.8 Å) the R-merge value was greater than 10.6 % yet decreased to an acceptable value at 4.0 Å. The unit cell dimensions were slightly smaller under Cryostream (98.1 Å x 98.1 Å x 240.1 Å) thereby preventing these data being merged with data previously acquired. The systematic absences in the reflections again indicated a $P4_12_12$ or $P4_32_12$ space group.

8.4 DISCUSSION

The preliminary crystallographic analysis of the *P. furiosus* citrate synthase indicated that the enzyme crystallised readily at room temperature. This may indicate the rigidity of the molecule at room temperature. From the preliminary crystallographic analysis, it was observed that the crystals generated in different conditions had the same unit cell dimensions and space group. Under Cryostream, the crystal was more resistant to deterioration in the X-ray beam and allowed the collection of 600 frames of data to a resolution of 4.0 Å. However, shrinkage of the unit cell prevented these data being merged with data collected at room temperature; unit cell shrinkage is a common feature under Cryostream.

The fundamental problem in interpreting crystallographic data is that the intensities and directions of the X-rays can be calculated directly from the data collected from the native enzyme but the phase of the diffracted rays is not available. To approach this problem, attempts are now being made (R. J. M. Russell, University of Bath) to construct heavy atom derivatives of the citrate synthase crystals and also to solve the structure by molecular replacement. The latter technique uses the *T. acidophilum* structure or the homology model of the *P. furiosus* citrate synthase sequence as a search model for the new structure. This is then used to find the position and orientation of the *P. furiosus* citrate synthase within the unit cell. The phases from the search model can be used to commence the refinement process for the new structure. Initial attempts to observe overlap between the search model and the 4.0 Å crystallographic data generated from the *P. furiosus* crystals have proved unsuccessful. It is possible that the degree of domain closure between the large and small domains of each monomer in the *P. furiosus* enzyme is greater than in the *T. acidophilum* enzyme (the degree of domain closure in the *T. acidophilum* enzyme is greater than in the pig enzyme (Section 1.1.4)). Therefore, the data are now being processed using just one domain of the *T. acidophilum* enzyme as a search model.

CHAPTER 9

PERSPECTIVE AND FUTURE WORK

It has become apparent that the stability of enzymes from hyperthermophilic sources is determined by a balance of intrinsic and extrinsic factors. The intrinsic factors, those derived entirely from the amino acid sequence, stabilise the folded conformation of the enzyme by enhancing the non-covalent interactions over those observed in non-thermostable enzymes. This results in a compact enzyme that is rigid at room temperature yet exhibits a similar degree of flexibility and catalytic activity at its optimum temperature to that of its non-thermostable counterpart at the lower temperature. The thermostable enzyme may also have adapted to avoid amino acids whose side chains are susceptible to chemical modification at high temperatures.

In addition to intrinsic factors, the intracellular enzymes of a hyperthermostable enzyme may also be stabilised by compatible solutes present in the cell cytoplasm. By this, intracellular enzymes are intrinsically adapted to the lower end of the growth temperature range yet can function at the higher end by the controlled synthesis of compatible solutes. This property must bestow a selective advantage to the organism in its natural habitat as the temperature of its immediate environment is likely to fluctuate. Compatible solutes enable the enzymes to be catalytically efficient over a wide temperature range, thus increasing the range over which the organism can survive.

This thesis describes the cloning, sequencing and expression in *E. coli* JM105 of the gene encoding citrate synthase from the hyperthermophilic Archaeon, *Pyrococcus furiosus*. The recombinant enzyme was functional, hyperthermostable and crystallised readily, to enable a preliminary crystallographic analysis. Thus, a convenient system has been generated whereby the intrinsic and extrinsic contributions to protein thermostability can be investigated.

Definition of the intrinsic factors that contribute to protein thermostability requires detailed scrutiny of the crystal structure of the hyperthermostable protein and comparison to the crystal structure of non-thermostable counterparts. The progress made towards this is described in Section 8.4. This has now extended to the use of synchrotron radiation sources for the collection of high resolution crystal data for the *P. furiosus* citrate synthase (R. J. M. Russell, University of Bath). Investigation of the contribution of extrinsic factors to protein stability is facilitated by the simple spectrophotometric assay for citrate synthase. Current research is now focussing on the effects of protein concentration, buffer composition and compatible solutes on the thermostability of this enzyme (N. Tolliday, University of Bath).

The work reported in this thesis complements and extends the study of the moderately-thermostable citrate synthase of *T. acidophilum*. Structural features conferring thermostability have been predicted from the crystal structure of this enzyme and are now being engineered into the pig heart enzyme by site directed mutagenesis (M. M^{c}Cormack, University of Bath). A similar situation can be envisaged for the testing of hyperthermostabilising features derived from the *P. furiosus* crystal structure within the *T. acidophilum* enzyme. This may be advantageous as the phylogenetic relationship between the two archaeal organisms is closer than the relationship between the *T. acidophilum* and the pig enzyme.

The gene for the citrate synthase of the thermophilic Archaeon, *S. solfataricus*, which grows optimally at 87°C is currently being targetted for a similar investigation (G. Munro, University of Bath). This sequence may be especially interesting, as the upper temperature limit of enzyme stability corresponds to the temperature at which the contribution of the hydrophobic effect has been suggested to be optimal [as reviewed by Dill, 1990]. Thus, a gradation of features may be observed between the *T. acidophilum* and *S. solfataricus* enzyme that does not continue to the *P. furiosus* enzyme where alternative mechanisms might be operating. In order to pursue this line of investgation in combination with microbial phylogeny, it may also be interesting to target the citrate synthase of a thermophilic Bacterium such as *Thermotoga maritima* or *Aquifex pyrophilus* which grow optimally at 80°C and 85°C, respectively. Features that are shared only by the thermophilic citrate synthases of *P. furiosus* (or *S. solfataricus*) and *Thermotoga maritima* (or *Aquifex pyrophilus*), despite the phylogenetic distance between the species, may indicate features involved solely in the adaptation of that enzyme to the extreme environment.

The ultimate goal of this programme, which uses citrate synthase as a model protein, is to elucidate general features that confer thermostability to the protein. It is probable that general features will be observed, but the mechanism by which they are accommodated within the protein structure will be enzyme-specific. Thus, targetted mutagenesis, to engineer thermostability into a protein, will require detailed knowledge of the protein structure. The work described in this thesis has provided the potential for extending the citrate synthase crystal structure comparison while generating ideas regarding the thermostabilisation of this enzyme. These will serve as a starting point for future investigations.

REFERENCES

Achenbach-Richter, L., Gupta, R., Zillig, W., Woese, C. R. (1988) 'Rooting the archaebacterial tree - the pivotal role of *Thermococcus celer* in archaebacterial evolution', *System. Appl. Microbiol.* **10**, 231-240

Ahern, T. J., Klibanov, A. M. (1985) 'The mechanism of irreversible enzyme inactivation at 100°C', *Science* 228, 1280-1284

Alefounder, P. R., Perham, M. (1989) 'Identification, molecular cloning and sequence analysis of a gene cluster encoding the class II fructose-1,6-bisphosphate aldolase, 3-phosphoglycerate kinase and a putative second glyceraldehyde-3-phosphate dehydrogenase of *Escherichia coli*', *Mol. Microbiol.* **3**, 723-732

Alter, G. M., Casazza, J. P., Zhi, W., Nemeth, P., Srere, P. A., Evans, C.T. (1990) 'Mutation of essential catalytic residues in pig citrate synthase', *Biochemistry* **29**, 7557-7563

Anderson, S. C. K., Powrie, R., Mitchell, C. G. (1993) '*Pseudomonas* citrate synthase; purification and characterisation', *Biochem. Soc. Trans.* **22**, 41S

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K. (Editors) (1987) 'Current protocols in molecular biology', Wiley Interscience, USA

Barns, S. M., Fundyga, R. E., Jeffries, M. W., Pace, N. R. (1994) 'Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment', *Proc. Natl. Acad. Sci. USA* **91**, 1609-1613

Barton, G. J., Sternberg, M. J. E. (1990) 'Flexible protein sequence patterns - a sensitive method to detect weak structural similarities', *J. Mol. Biol.* **212**, 389-402

Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A., Matthews, B. W. (1992) 'Dissection of helix capping in T4 lysozyme by structural and thermodynamic analysis of six amino acid substitutions at Thr 59', *Biochemistry* **31**, 3590-3596

Blake, P. R., Park, J. B., Bryant, F. O., Aono, S., Magnuson, J. K., Eccleston, E., Howard, J. B., Summers, M. F., Adams, M. W. W. (1991) 'Determinants of protein hyperthermostability: purification and amino acid sequence of rubredoxin from the hyperthermophilic Archaebacterium, *Pyrococcus furiosus* and secondary structure of the zinc adduct by NMR', *Biochemistry* **30**, 10885-10895

Blamey, J. M., Adams M. W. W. (1993) 'Purification and characterization of pyruvate ferredoxin oxido-reductase from the hyperthermophilic Archaeon, *Pyrococcus furiosus*', *Biochim. Biophys. Acta.* **1161**, 19-27

Bloch, W. (1991) 'A biochemical perspective of the polymerase chain reaction', *Biochemistry* **30**, 2735-2747

Blumenthals, I. L., Robinson, A. S., Kelly, R. M. (1990) 'Characterization of sodium dodecyl sulfate-resistant proteolytic activity in the hyperthermophilic Archaebacterium, *Pyrococcus furiosus*', *Appl. Environ. Microbiol.* **56**, 1992-1998

Bradford, M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal. Biochem.* **72**, 248-254

Brünger, A. T. (1990) XPLOR Manual Version 2.1, Yale University, CT., USA

Bryant, F. O., Adams M. W. W. (1989) 'Characterization of hydrogenase from the hyperthermophilic Archaebacterium, *Pyrococcus furiosus*', *J. Biol. Chem.* **264**, 5070-5079

Burley, S. K., Petsko, G. A. (1985) 'Aromatic-aromatic interaction: a mechanism of protein structure stabilization', *Science* **229**, 23-28

Consalvi, V., Chiaraluce, R., Politi, L., Vaccaro, R., De Rosa, M., Scandurra, R. (1991) 'Extremely thermostable glutamate dehydrogenase from the hyperthermophilic Archaebacterium, *Pyrococcus furiosus*', *Eur. J. Biochem.* **202**, 1189-1196

Constantino, H. R., Brown, S. H., Kelly, R. M. (1990) 'Purification and characterization of an α-glucosidase from a hyperthermophilic Archaebacterium, *Pyrococcus furiosus*, exhibiting a temperature optimum of 105 to 115°C', *J. Bacteriol.* **172**, 3654-3660

Creti, R., Citarella, F., Tiboni, O., Sanangelantoni, A., Palm, P., Cammarano, P. (1991) 'Nucleotide sequence of a DNA region comprising the gene for elongation factor 1α (EF-1α) from the ultrathermophilic Archaeote, *Pyrococcus woesei*: phylogenetic implications', *J. Mol. Evol.* **33**, 332-342

Danson, M.J. (1988) 'Archaebacteria: the comparative enzymology of their central metabolic pathways', *Adv. Microb. Phys.* **29**, 166-223

Danson, M. J. (1993) 'Central metabolism of the Archaea' in 'The Biochemistry of the Archaea' edited by Kates M., Kushner, D. J. and Matheson A. T. [Elsevier Science Publishers, Amsterdam]

Dao-pin, S., Sauer, U., Nicholson, H., Matthews, B. W. (1991) 'Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis', *Biochemistry* **30**, 7142-7153

David, M., Lubinsky-Mink, S., Ben-Zvi, A., Suissa, M., Ulitzur, S., Kuhn, J. (1991) 'Citrate synthase from *Mycobacterium smegmatis*', *Biochem. J.* **278**, 225-234

Davies, G. J., Gamblin, S. J., Littlechild, J. A., Watson, H. C. (1993) 'The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent', *Proteins* **15**, 283-289

Delong, E. F., Wu, K. Y., Prézelin, B. B., Jovine, R. V. M. (1994) 'High abundance of Archaea in Antarctic marine picoplankton', *Nature* **371**, 695-697

Devereux, J., Haeberli, P., Smithies, O. (1984) 'A comprehensive set of sequence analysis programs for the VAX', *Nucleic Acids Res.* **12**, 387-395

Dill, K. A. (1990) 'Dominant forces in protein folding', *Biochemistry* **29**, 7133-7155

Donald, L. J., Duckworth, H. W. (1987) ' Expression and base sequence of the citrate synthase gene of *Acinetobacter anitratum*', *Biochem. Cell. Biol.*, **65**, 930-938

Donald, L. J., Molgat, G. F., Duckworth, H. W. (1989) 'Cloning, sequencing and expression of the gene for NADH-sensitive citrate synthase of *Pseudomonas aeruginosa*', *J. Bacteriol.* **171**, 5542-5550

Dower, W. J., Miller, J. F., Ragsdale, C. W. (1988) 'High efficiency transformation of *Escherichia coli* by high voltage electroporation', *Nucleic Acids Res.* **16**, 6127-6145

Eggen, R. I. L., Geerling, A. C. M., Watts, J., de Vos, W. M. (1990) 'Characterization of pyrolysin, a hyperthermoactive serine protease from the Archaebacterium, *Pyrococcus furiosus*', *FEMS Microbiol. Lett.* **71**, 17-20

Eggen, R. I. L., Geerling, A. C. M., Waldkotter, K., Antranikian, G., de Vos, W. M. (1993) 'The glutamate dehydrogenase-encoding gene of the hyperthermophilic Archaeon, *Pyrococcus furiosus* - sequence, transcription and analysis of the deduced amino acid sequence', *Gene* **132**, 143-148

Eisenthal, R., Cornish-Bowden, A. (1974) 'The direct linear plot', *Biochem. J.* **175**, 715-720

Erauso, G., Reysenbach, A. L., Godfroy, A., Meunier, J. R., Crump, B., Partensky, F., Baross, J. A., Marteinsson, V., Barbier, G., Pace, N. R., Prieur, D. (1993) '*Pyrococcus abyssi* sp. nov., a new hyperthermophilic Archaeon isolated from a deep-sea hydrothermal vent', *Arch. Microbiol.* **160**, 338-349

Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., Matthews, B. W. (1992) 'Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect', *Science* **255**, 178-183

Evans, C. T., Owens, D. D., Sumegi, B., Kispal, G., Srere, P. A. (1988) 'Isolation, nucleotide sequence, and expression of a cDNA encoding pig citrate synthase', *Biochemistry* **27**, 4680-4686

Fiala, G., Stetter, K. O. (1986) '*Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic Archaebacteria growing optimally at 100°C', *Arch. Microbiol.* **145**, 56-61

Felsenstein, J. (1988) 'Phylogenies from molecular sequences: inferences and reliability', *Ann. Rev. Genet.* **22**, 521-565

Felsenstein, J. (1993) 'PHYLIP (Phylogeny Inference Package) Version 3.5c - User manual (University of Washington)

Fersht, A. R., Serrano, L. (1993) 'Principles of protein stability derived from protein engineering experiments', *Curr. Opinion Struct. Biol.* **3**, 75-83

Fukaya, M., Takemura, H., Okumura, H., Kawamura, Y., Horinouchi, S., Beppu, T. (1990) 'Cloning of genes responsible for acetic acid resistance in *Acetobacter acetii*', *J. Bacteriol.* **172**, 2096-2104

Geiger, T., Clarke, S. (1987) 'Deamidation, isomerization and racemization at asparaginyl and aspartyl residues in peptides', *J. Biol. Chem.* **262**, 785-794

Gyllensten, U. B., Erlich, H. A. (1988) 'Generation of single stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA-DQA* locus', *Proc. Natl. Acad. Sci. USA* **85**, 7652-7656

Hanahan, D. (1983) 'Studies on the transformation of *Escherichia coli* with plasmids', *J. Mol. Biol.* **166**, 557-580

Hanahan, D. (1985) 'Techniques for the transformation of *E. coli*', in DNA Cloning, Volume 1 - a practical approach edited by Glover, D. M. [IRL Press, Oxford]

Heinzen, R. A., Frazier, M. E., Mallavia, L. P. (1991) 'Sequence and linkage analysis of the *Coxiella burnetii* citrate synthase-encoding gene', *Gene* **109**, 63-69

Hensel, R., Jakob, I. (1994) 'Stability of glyceraldehyde-3-phosphate dehydrogenase from hyperthermo-philic Archaea at high temperature', *System. Appl. Microbiol.* **16**, 742-745

Hensel, R., König, H. (1988) 'Thermoadaptation of methanogenic bacteria by intracellular ion concentration', *FEMS Microbiol. Lett.* **49**, 75-79

Hensel, R., Kohlhoff, M., Hess, D. (1993) 'Factors influencing the thermostability of enzymes from hyperthermophilic Archaea', in Proceedings of Thermophiles '93: An International Conference on the Science and Technology of Thermophiles (Hamilton, New Zealand) A11

Hol, W. C. J. (1985) 'The role of the α-helix dipole in protein function and structure', *Prog. Biophys. Molec. Biol.* **45**, 149-195

Horovitz, A., Matthews, J. M., Fersht, A. R. (1992) 'α-Helix stability in proteins II. Factors that influence stability at an internal position', *J. Mol. Biol.* **227**, 560-568

Huber, R., Kristjansson, J. K., Stetter, K. O. (1987) '*Pyrobaculum* gen. nov., a new genus of neutrophilic rod-shaped Archaebacteria from continental solfataras growing optimally at 100°C', *Arch. Microbiol.* **149**, 95-101

Huber, R., Kurr, M., Jannasch, H. W., Stetter, K. O. (1989) 'A novel group of abyssal methanogenic Archaebacteria (*Methanopyrus*) growing at 110°C', *Nature* **342**, 833-834

Hutchinson, E. G., Thornton, J. M. (1990) 'HERA - a program to draw schematic diagrams of protein secondary structure', *Proteins* **8**, 203-212

Ishikawa, K., Nakamura, H., Morikawa, K., Kanaya, S. (1993) 'Stabilization of *Escherichia coli* ribonuclease HI by cavity-filling mutations within a hydrophobic core', *Biochemistry* **32**, 6171-6178

Itoh, T. (1989) 'Sequence-analysis of the peptide-elongation factor EF-2 gene, downstream from those of ribosomal-proteins H-S12 and H-S7, from the archaebacterial extreme halophile, *Halobacterium halobium*', *Eur. J. Biochem.* **186**, 213-219

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., Miyata, T. (1989) 'Evolutionary relationship of Archaebacteria, Eubacteria, and Eukaryotes inferred from phylogenetic trees of duplicated genes', *Proc. Natl. Acad. Sci. USA* **86**, 9355-9359

Jaenicke, R. (1991) 'Protein stability and molecular adaptation to extreme conditions', *Eur. J. Biochem.* **202**, 715-728

James, K. D., Bonete, M. J., Byrom, D., Danson, M. J., Hough, D. W. (1991) 'Citrate synthase from *Haloferax volcanii*: enzyme purification and gene cloning', *Biochem. Soc. Trans.* **20**, 12S

Jin, S., Sonenshein, A. L. (1994) 'Identification of two distinct *Bacillus subtilis* citrate synthase genes', *J. Bacteriol.* **176**, 4669-4679

Jones, T. A., Zou, J. Y., Cowan, S. W., Kjeldgaard, M. (1991) 'Improved methods for building protein models in electron density maps and the location of errors in these models', *Acta Crystall.* **A47**, 110-119

Kabsch, W., Sander, C. (1983) 'Dictionary of protein secondary structure - pattern recognition of hydrogen bonded and geometrical features', *Biopolymers* **22**, 2577-2637

Kabsch, W. (1988) 'Automatic-indexing of rotation diffraction patterns', *J. Appl. Crystall.* **21**, 67-71

Karpusas, M., Branchaud, B., Remington, S. J. (1990) 'Proposed mechanism for the condensation reaction of citrate synthase: 1.9 Å structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A', *Biochemistry* **29**, 2213-2219

Karpusas, M., Holland, D., Remington, S. J. (1991) '1.9 Å structures of ternary complexes of citrate synthase with D- and L-malate: mechanistic implications', *Biochemistry* **30**, 6024-6031

Kellis, J. T., Nyberg, K., Šali, D., Fersht, A. R. (1988) 'Contribution of hydrophobic interactions to protein stability', *Nature* **333**, 784-786

Kengen, S. W., de Bok, F. A. M., van Loo, N. D., Dijkema, C., Stams A. J. M., de Vos, W. M. (1994) 'Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*', *J. Biol. Chem.* **269**, 17537-17541

Keohavong, P., Thilly, W. G. (1989) 'Fidelity of DNA polymerases in DNA amplification', *Proc. Natl. Acad. Sci. USA* **86**, 9253-9257

Keohavong, P., Ling, L. Dias, C., Thilly, W. G. (1993) 'Predominant mutations induced by the *Thermococcus litoralis*, Vent DNA polymerase during DNA amplification *in vitro*', *PCR Methods and Applications* **2**, 288-292

Kjems, J., Larsen, N., Dalgaard, J. Z., Garrett, R. A., Stetter, K. O. (1992) 'Phylogenetic relationships amongst the hyperthermophilic Archaea determined from partial 23S rRNA gene sequences', *System. Appl. Microbiol.* **15**, 203-208

Koch, R., Zablowski, P., Spreinat, A., Antranikian, G. (1990) 'Extremely thermostable amylolytic enzyme from the Archaebacterium, *Pyrococcus furiosus*', *FEMS Microbiol. Lett.* **71**, 21-26

Kong, H. M., Kucera, R. B., Jack, W. E. (1993) 'Characterization of a DNA polymerase from the hyperthermophilic Archaeon, *Thermococcus litoralis* - Vent DNA polymerase, steady-state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities', *J. Biol. Chem.* **268**, 1965-1975

Laemmli, U. K. (1970) 'Cleavage of structural proteins during the assembly of the head of Bacteriophage T4', *Nature* **227**, 680-685

Lake, J. A. (1988) 'Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences', *Nature* **331**, 184-186

Lanyi, J. K. (1974) 'Salt-dependent properties of proteins from extremely halophilic bacteria', *Bacteriol. Rev.* **38**, 272-290

Lechner, K., Heller, G., Böck, A. (1988) 'Gene for the diphtheria toxin-susceptible elongation factor II from *Methanococcus vannielii*', *Nucleic Acids Res.* **16**, 7817-7826

Lill, U., Lefrank, S., Henschen, A., Eggerer, H. (1992) 'Conversion, by limited proteolysis, of an archaebacterial citrate synthase into essentially, a citryl-CoA hydrolase', *Eur. J. Biochem.* **208**, 459-466

Loewenthal, R., Sancho, J., Fersht, A. R. (1992) 'Histidine-aromatic interactions in barnase', *J. Mol. Biol.* **224**, 759-770

Lozano, P., Combes, D., Iborra, J. L. (1994) 'Effect of polyols on α-chymotrypsin thermostability: a mechanistic analysis of enzyme stabilization', *J. Biotech.* **35**, 9-18

Lundberg, K. S., Shoemaker, D. D., Short, J. M., Sorge, J. A., Adams, M. W. W., Mathur, E. (1991) 'High fidelity amplification with a thermostable DNA polymerase isolated from *Pyrococcus furiosus*', *Gene* **108**, 1-6

Menéndez-Arias, L., Argos, P. (1989) 'Engineering protein thermal stability: sequence statistics point to residue substitutions in α-helices', *J. Mol. Biol.* **206**, 397-406

Morris, A. L., Macarthur, M. W., Hutchinson, E. G., Thornton, J. M. (1992) 'Stereochemical quality of protein structure co-ordinates', *Proteins* **12**, 345-364

Mukund, S., Adams M. W. W. (1991) 'The novel tungsten-iron-sulphur protein of the hyperthermophilic archaebacterium, *Pyrococcus furiosus*, is an aldehyde ferredoxin oxidoreductase', *J. Biol. Chem.* **266**, 14208-14216

Mullis, K., Faloona, F. A. (1987) 'Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction', *Methods Enzymol.* **155**, 335-350

Mullis, K., Faloona, F. A., Scharf, S., Saiki, R., Horn, G., Erlich, H. A. (1986) 'Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction', *Cold Spring Harbor Symp. Quant. Biol.* **51**, 263-273

Ner, S. S., Bhayana, V., Bell, A. W., Giles I. G., Duckworth, H. W., Bloxham, D. P. (1983) 'Complete sequence of the *gltA* gene encoding citrate synthase in *Escherichia coli*', *Biochemistry* **22**, 5243-5249

Numata, O., Takemasa, T., Takagi, I., Hirono, M., Hirano, H., Chiba, J., Watanabe, Y. (1991) '*Tetrahymena* 14 nm filament-forming protein has citrate synthase activity', *Biochem. Biophys. Res. Commun.* **174**, 1028-1034

Ochman, H., Gerber, A. S., Hartl, D. L. (1988) 'Genetic applications of an inverse polymerase chain reaction', *Genetics* **120**, 621-625

Olsen, G. J., Woese, C. R. (1993) 'Ribosomal RNA: a key to phylogeny', *FASEB J.* **7**, 113-123

Patton, A. J., Hough, D. W., Towner, P., Danson, M. J. (1993) 'Does *E. coli* possess a second citrate synthase gene?', *Eur. J. Biochem.* **215**, 75-81

Pauptit, R. A., Karlsson, R., Picot, D., Jenkins, J. A., Niklaus-Reimer, A. S., Jansonius, J. N. (1988) 'Crystal structure of neutral protease from *Bacillus cereus* refined at 3.0 Å resolution and comparison with the homologous but more thermostable enzyme, thermolysin', *J. Mol. Biol.* **199**, 525-537

Peak, M. J., Peak, J. G., Stevens, F. J., Blamey, J., Mai, X., Zhou, Z. H., Adams, M. W. W. (1994) 'The hyperthermophilic glycolytic enzyme enolase in the Archaeon, *Pyrococcus furiosus*: comparison with mesophilic enolases', *Arch. Biochem. Biophys.* **313**, 280-286

Pechmann, H., Tesch, A., Klink, F. (1991) 'Cloning and sequencing of the *fus*-gene encoding elongation factor 2 in the Archaebacterium, *Thermoplasma acidophilum*', *FEMS Microbiol. Lett.* **79**, 51-56

Qaw, F. S., Brewer, J. M. (1986) 'Arginyl residues and thermal stability in proteins', *Mol. Cell. Biochem.* **71**, 121-127

Remington, S. J. (1992) 'Structure and mechanism of citrate synthase', *Curr. Top. Cell. Regul.* **33**, 209-219

Remington, S. J., Wiegand, G., Huber, R. (1982) ' Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 Å and 1.7 Å resolution', *J. Mol. Biol.* **158**, 111-152

Robb, F. T., Park, J. B., Adams, M. W. W. (1992) 'Characterization of an extremely thermostable glutamate dehydrogenase: a key enzyme in the primary metabolism of the hyperthermophilic Archaebacterium, *Pyrococcus furiosus*' *Biochim. Biophys. Acta.* **1120**, 267-272

Rosekrantz, M., Alam, T., Kim, K., Clark, B. J., Srere, P. A., Guarente, L.P. (1986) 'Mitochondrial and non-mitochondrial citrate synthases in *Saccharomyces cerevisiae* are encoded by distinct genes.', *Mol. Cell. Biol.* **6**, 4509-4515

Russell, R. J. M. (1994) Ph.D thesis, University of Bath

Russell, R. J. M., Hough, D. W., Danson, M. J., Taylor, G. L. (1994) 'The crystal structure of citrate synthase from the thermophilic Archaeon, *Thermoplasma acidophilum*', *Structure* (in press)

Saiki, R. K., Scharf, F., Faloona, K. B., Mullis, K. B., Horn G. T., Erlich, H. A., Arnheim, N. (1985) 'Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia', *Science* **230**, 1350-1354

Šali, D., Bycroft, M., Fersht, A. R., (1991) 'Surface electrostatic interactions contribute little to stability of barnase', *J. Mol. Biol.* **220**, 779-788

Sambrook, J., Fritsch, E. F., Maniatis, T. (1989) 'Molecular cloning - a laboratory manual', Cold Spring Harbor Laboratory Press, Cold Spring Laboratory, New York

Sanger, F., Niklen, S., Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors', *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467

Schäfer, T., Schönheit, P. (1991) 'Pyruvate metabolism of the hyperthermophilic Archaebacterium, *Pyrococcus furiosus*', *Arch. Microbiol.* **155**, 366-377

Schäfer, T., Schönheit, P. (1992) 'Maltose fermentation to acetate, CO_2 and H_2 in the anaerobic, hyperthermophilic Archaeon, *Pyrococcus furiosus*: evidence for the operation of a novel sugar fermentation pathway', *Arch. Microbiol.* **158**, 188-202

Schäfer, T., Schönheit, P. (1993) 'Gluconeogenesis from pyruvate in the hyperthermophilic Archaeon, *Pyrococcus furiosus*: involvement of reactions of the Embden-Meyerhof pathway', *Arch. Microbiol.* **159**, 354-363

Schendel, F. J., August, P. R., Anderson, C. R., Hanson, R. S., Flickinger, M. C. (1992) 'Cloning and nucleotide sequence of the gene coding for citrate synthase from a thermotolerant *Bacillus* sp.' *Appl. Environ. Microbiol.* **58**, 335-345

Scholz, S., Sonnenbichler, J., Schäfer, W., Hensel, R. (1992) ' Di-myo-inositol-1,1'-phosphate: a new inositol phosphate isolated from *Pyrococcus woesei*', *FEBS Lett.* **2/3**, 239-242

Schröder, J., Klink, F. (1991) 'Gene for the ADP-ribosylatable elongation factor 2 from the extreme thermoacidophilic Archaebacterium, *Sulfolobus acidocaldarius*', *Eur. J. Biochem.* **195**, 321-327

Schultes, V., Deutzmann, R., Jaenicke, R. (1990) 'Complete amino acid sequence of glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic Eubacterium, *Thermotoga maritima*', *Eur. J. Biochem.* **192**, 25-31

Serrano, L., Horovitz, A., Avron, B., Bycroft, M., Fersht, A. R. (1990) 'Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double mutant cycles', *Biochemistry* **29**, 9343-9352

Serrano, L., Bycroft, M., Fersht, A. R. (1991) 'Aromatic-aromatic interactions and protein stability', *J. Mol. Biol.* **218**, 465-475

Serrano, L., Neira, J. L., Sancho, J., Fersht, A. R. (1992a) 'Effect of alanine versus glycine in α-helices on protein stability', *Nature* **356**, 453-455

Serrano, L., Sancho, J., Hirshberg, M., Fersht, A. R. (1992b) 'α-Helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces', *J. Mol. Biol.* **227**, 544-559

Shepherd, N. (1993) Final year project report, University of Bath

Shirley, B. A., Stanssens, P., Hahn, U., Pace, C. N. (1992) 'Contribution of hydrogen bonding to the conformational stability of ribonuclease T1', *Biochemistry* **31**, 725-732

Shortle, D., Stites, W. E., Meeker, A. K. (1990) 'Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease', *Biochemistry* **29**, 8033-8041

Silver, J., Keerikatte, V. (1989) 'Novel use of the polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus', *J. Virol.* **63**, 1924-1928

Singh, J., Thornton, J. M. (1985) 'The interaction between phenylalanine rings in proteins', *FEBS Lett.* **191**, 1-6

Snowden, L. J., Blumenthals, I. I., Kelly, R. M. (1992) 'Regulation of proteolytic activity in the hyperthermophile, *Pyrococcus furiosus*' *Appl. Environ. Microbiol.* **58**, 1134-1141

Southern, E. (1975) 'Detection of specific sequences among DNA fragments separated by gel electrophoresis', *J. Mol. Biol.* **98**, 503-517

Srere, P. A., Brazil, H., Gonen, L. (1963) 'The citrate condensing enzyme of pigeon breast muscle and moth flight muscle', *Acta. Chem. Scand.* **17**, S129-S134

Stanier, R. Y., van Niel, C. B. (1962) 'The concept of a bacterium', *Arch.Mikrobiol* **42**, 17-35

Stephenson, R. C., Clarke, S. (1989) 'Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins', *J. Biol. Chem.* **264**, 6164-6170

Stetter, K. O., König, H., Stackebrandt, E. (1983) '*Pyrodictium* gen. nov., a new genus of submarine disc-shaped sulfur reducing Archaebacteria growing optimally at 105°C', *System. Appl. Microbiol.* **4**, 535-551

Sutherland, K. J., Henneke, C. M., Towner, P., Hough, D. W., Danson M. J. (1990) 'Citrate synthase from the thermophilic Archaebacterium, *Thermoplasma acidophilum*', *Eur. J. Biochem.* **194**, 839-844

Tautz, D., Renz, M. (1983) 'An optimised freeze-squeeze method for the recovery of DNA fragments from agarose gels', *Anal. Biochem.* **132**, 14-19

Tindall, K. R., Kunkel, T. A. (1988) 'Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase', *Biochemistry* **27**, 6008-6013

Triglia, T., Peterson, M. G., Kemp, D. J. (1988) 'A procedure for an *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences', *Nucleic Acids Res.* **16**, 8186

Tsou, C. (1986) 'Location of the active sites of some enzymes in limited and flexible molecular regions', *TIBS* **11**, 427-429

Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., Itakura, K. (1979) 'Hybridization of synthetic oligodeoxyribonucleotides to ϕ χ 174 DNA: the effect of a single base pair mismatch', *Nucleic Acids Res.* **6**, 3543-3557

Wheelis, M. L., Kandler, O., Woese, C.R. (1992) 'On the nature of global classification', *Proc. Natl. Acad. Sci. USA* **89**, 2930-2934

Woese, C. R., Fox G. E. (1977) 'Phylogenetic structure of the prokaryotic domain: the primary kingdoms', *Proc. Natl. Acad. Sci. USA* **74**, 5088-5090

Woese, C. R., Kandler, O., Wheelis, M. L. (1990) 'Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya', *Proc. Natl. Acad. Sci. USA* **87**, 4576-4579

Wood, D. O., Williamson, L. R., Winkler, H. H., Krause, D. L. (1987) 'Nucleotide sequence of the *Rickettsia prowazekii* citrate synthase gene', *J. Bacteriol.* **169**, 3564-3572

Zaccai, G., Eisenberg, H. (1990) 'Halophilic proteins and the influence of solvent on protein stabilization', *TIBS* **15**, 333-337

Zale, S. E., Klibanov, A. M. (1986) 'Why does ribonuclease irreversibly inactivate at high temperatures?', *Biochemistry* **25**, 5432-5444

Zhi, W., Srere, P. A., Evans, C. T. (1991) 'Conformational stability of pig citrate synthase and some active site mutants', *Biochemistry* **30**, 9281-9286

Zillig, W., Holz, L., Janekovic, D., Klenk, H. P., Insel, E., Trent, J., Wunderl, S., Forjaz, V. H., Coutinho, R., Ferreira, T. (1990) '*Hyperthermus butylicus*, a hyperthermophilic, sulfur-reducing Archaebacterium that ferments peptides', *J. Bacteriol.* **172**, 3959-3965

Zillig, W., Palm, P., Rieter, W. D., Gropp, F., Pühler, G., Klenk, H. P. (1988) 'Comparative evaluation of gene expression', *Eur. J. Biochem.* **173**, 473-482

Zwickl, P., Fabry, S., Bogedain, C., Haas, A., Hensel, R. (1990) 'Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic Archaebacterium, *Pyrococcus woesei*: characterization of the enzyme, cloning and sequencing of the gene and expression in *Escherichia coli*', *J. Bacteriol.* **172**, 4329-4338